1. Older Tickets
    1. The Fisher-Neyman Factorization Theorem
    2. Gelfand n-widths
    3. Molecular Structure and Physical Properties
    4. The Environmental Impact of the Manufacturing of Seminconductors (ticket #5622)
2. Newer Tickets
    1. CNXML 0.5 Stress Test
    2. Frequency Domain Problems (ticket #6364)
    3. Energy in a mechanical wave
    4. Árvores binárias de pesquisa
    5. Hypothesis Testing of Single Mean and Single Proportion: Examples (ticket #6053)
    6. Rational Functions (ticket #5775)
    7. Astronomical Image Deconvolution: Image Processing with Weiner Filter
    8. Form in Music (ticket #5317)
    9. Autocorrelation of Random Processes
    10. State-Variable Representation of Discrete-Time Systems
    11. Introduction
    12. Fast Convolution
    13. Acquiring a Signal VI
    14. Relative Atomic Masses and Empirical Formulae

The Fisher-Neyman Factorization Theorem

Determining a sufficient statistic directly from the definition can be a tedious process. The following result can simplify this process by allowing one to spot a sufficient statistic directly from the functional form of the density or mass function.

Fisher-Neyman Factorization Theorem

Let $\mathrm{f}_{\boldsymbol{\theta}}(\boldsymbol{x})$ be the density or mass function for the random vector $\boldsymbol{x}$, parametrized by the vector $\boldsymbol{\theta}$. The statistic $\boldsymbol{t} = T(\boldsymbol{x})$ is sufficient for $\boldsymbol{\theta}$ if and only if there exist functions $a(\boldsymbol{x})$ (not depending on $\boldsymbol{\theta}$) and $\mathrm{b}_{\boldsymbol{\theta}}(\boldsymbol{t})$ such that

$$\mathrm{f}_{\boldsymbol{\theta}}(\boldsymbol{x}) = a(\boldsymbol{x})\,\mathrm{b}_{\boldsymbol{\theta}}(\boldsymbol{t})$$

for all possible values of $\boldsymbol{x}$.

In an earlier example we computed a sufficient statistic for a binary communication source (independent Bernoulli trials) from the definition. Using the above result, this task becomes substantially easier.

Example:

Bernoulli Trials Revisited

Suppose $x_n \sim \mathrm{Bernoulli}(\theta)$ are IID, $\forall n, n = 1, \ldots, N : (n = 1, \ldots, N)$. Denote $\boldsymbol{x} = (x_1 \ldots x_n)^T$. Then

Equation:

$$\begin{aligned} \mathrm{f}_{\theta}(\boldsymbol{x}) &= \prod_{n=1}^{N} \theta^{x_n}(1-\theta)^{1-x_n} \\ &= \theta^{k}(1-\theta)^{N-k} \\ &= a(\boldsymbol{x})\,\mathrm{b}_{\theta}(k) \end{aligned}$$

where $k = \sum_{n=1}^{N} x_n$, $a(\boldsymbol{x}) = 1$, and $\mathrm{b}_\theta\left(k\right) = \theta^k (1-\theta)^{N-k}$. By the Fisher-Neyman factorization theorem, k is sufficient for θ.

The next example illustrates the appliction of the theorem to a continuous random variable.

Example:

Normal Data with Unknown Mean

Consider a normally distributed random sample $x_1, \ldots, x_N \sim \mathcal{N}(\theta, 1)$, IID, where θ is unknown. The joint pdf of $\boldsymbol{x} = (x_1 \ldots x_n)^T$ is

$$\mathrm{f}_\theta\left(\boldsymbol{x}\right) = \prod_{n=1}^{N} \mathrm{f}_\theta\left(x_n\right) = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} e^{\frac{-1}{2}\sum_{n=1}^{N}\left(x_n - \theta\right)^2}$$

We would like to rewrite $\mathrm{f}_\theta\left(\boldsymbol{x}\right)$ is the form of $a(\boldsymbol{x})\,\mathrm{b}_\theta\left(\boldsymbol{t}\right)$, where $\dim\left(\boldsymbol{t}\right) < N$. At this point we require a trick-one that is commonly used when manipulating normal densities, and worth remembering. Define $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$, the sample mean. Then

Equation:

$$\begin{aligned} \mathrm{f}_\theta\left(\boldsymbol{x}\right) &= \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} e^{\frac{-1}{2}\sum_{n=1}^{N}\left(x_n - \bar{x} + \bar{x} - \theta\right)^2} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} e^{\frac{-1}{2}\sum_{n=1}^{N}\left(x_n - \bar{x}\right)^2 + 2\left(x_n - \bar{x}\right)\left(\bar{x} - \theta\right) + \left(\bar{x} - \theta\right)^2} \end{aligned}$$

Now observe
Equation:

$$\begin{aligned}\sum_{n=1}^{N}\left(x_n-\bar{x}\right)\left(\bar{x}-\theta\right) &= \left(\bar{x}-\theta\right)\sum_{n=1}^{N}x_n-\bar{x}\\ &= \left(\bar{x}-\theta\right)\left(\bar{x}-\bar{x}\right)\\ &= 0\end{aligned}$$

so the middle term vanishes. We are left with

$$\mathrm{f}_{\theta}\left(\boldsymbol{x}\right)=\left(\frac{1}{2\pi}\right)^{\frac{N}{2}}e^{\frac{-1}{2}\sum_{n=1}^{N}\left(x_n-\bar{x}\right)^2}e^{\frac{-1}{2}\sum_{n=1}^{N}\left(\bar{x}-\theta\right)^2}$$

where $a(\boldsymbol{x})=\left(\frac{1}{2\pi}\right)^{\frac{N}{2}}e^{\frac{-1}{2}\sum_{n=1}^{N}\left(x_n-\bar{x}\right)^2}$,
$\mathrm{b}_{\theta}\left(\boldsymbol{t}\right)=e^{\frac{-1}{2}\sum_{n=1}^{N}\left(\bar{x}-\theta\right)^2}$, and $\boldsymbol{t}=\boldsymbol{x}$. Thus, the sample mean is a one-dimensional sufficient statistic for the mean.

Proof of Theorem

First, suppose $\boldsymbol{t}=T(\boldsymbol{x})$ is sufficient for $\boldsymbol{\theta}$. By definition, $\mathrm{f}_{\boldsymbol{\theta}|T(\boldsymbol{x})=\boldsymbol{t}}\left(\boldsymbol{x}\right)$ is independent of $\boldsymbol{\theta}$. Let $\mathrm{f}_{\boldsymbol{\theta}}\left(\boldsymbol{x},\boldsymbol{t}\right)$ denote the joint density or mass function for $(\boldsymbol{X},\boldsymbol{T}(\boldsymbol{X}))$. Observe $\mathrm{f}_{\boldsymbol{\theta}}\left(\boldsymbol{x}\right)=\mathrm{f}_{\boldsymbol{\theta}}\left(\boldsymbol{x},\boldsymbol{t}\right)$. Then
Equation:

$$\begin{aligned}\mathrm{f}_{\boldsymbol{\theta}}\left(\boldsymbol{x}\right) &= \mathrm{f}_{\boldsymbol{\theta}}\left(\boldsymbol{x},\boldsymbol{t}\right)\\ &= \mathrm{f}_{\boldsymbol{\theta}|\boldsymbol{t}}\left(\boldsymbol{x}\right)\mathrm{f}_{\boldsymbol{\theta}}\left(\boldsymbol{t}\right)\\ &= a(\boldsymbol{x})\,\mathrm{b}_{\boldsymbol{\theta}}\left(\boldsymbol{t}\right)\end{aligned}$$

where $a(\boldsymbol{x}) = \mathrm{f}_{\boldsymbol{\theta}|\boldsymbol{t}}\ (\boldsymbol{x})$ and $\mathrm{b}_{\boldsymbol{\theta}}\ (\boldsymbol{t}) = \mathrm{f}_{\boldsymbol{\theta}}\ (\boldsymbol{t})$. We prove the reverse implication for the discrete case only. The continuous case follows a similar argument, but requires a bit more technical work (Scharf, pp.82; Kay, pp.127).

Suppose the probability mass function for $\boldsymbol{x}$ can be written

$$\mathrm{f}_{\boldsymbol{\theta}}\ (\boldsymbol{x}) = a(\boldsymbol{x})\ \mathrm{b}_{\boldsymbol{\theta}}\ (\boldsymbol{x})$$

where $\boldsymbol{t} = T(\boldsymbol{x})$. The probability mass function for $\boldsymbol{t}$ is obtained by summing $\mathrm{f}_{\boldsymbol{\theta}}\ (\boldsymbol{x}, \boldsymbol{t})$ over all $\boldsymbol{x}$ such that $T(\boldsymbol{x}) = \boldsymbol{t}$:

Equation:

$$\begin{aligned}\mathrm{f}_{\boldsymbol{\theta}}\ (\boldsymbol{t}) &= \sum_{T(\boldsymbol{x})=\boldsymbol{t}} \mathrm{f}_{\boldsymbol{\theta}}\ (\boldsymbol{x}, \boldsymbol{t}) \\ &= \sum_{T(\boldsymbol{x})=\boldsymbol{t}} \mathrm{f}_{\boldsymbol{\theta}}\ (\boldsymbol{x}) \\ &= \sum_{T(\boldsymbol{x})=\boldsymbol{t}} a(\boldsymbol{x})\ \mathrm{b}_{\boldsymbol{\theta}}\ (\boldsymbol{t})\end{aligned}$$

Therefore, the conditional mass function of $\boldsymbol{x}$, given $\boldsymbol{t}$, is

Equation:

$$\begin{aligned}\mathrm{f}_{\boldsymbol{\theta}|\boldsymbol{t}}\ (\boldsymbol{x}) &= \frac{\mathrm{f}_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{t})}{\mathrm{f}_{\boldsymbol{\theta}}(\boldsymbol{t})} \\ &= \frac{\mathrm{f}_{\boldsymbol{\theta}}(\boldsymbol{x})}{\mathrm{f}_{\boldsymbol{\theta}}(\boldsymbol{t})} \\ &= \frac{a(\boldsymbol{x})}{\sum_{T(\boldsymbol{x})=\boldsymbol{t}} a(\boldsymbol{x})}\end{aligned}$$

This last expression does not depend on $\boldsymbol{\theta}$, so $\boldsymbol{t}$ is a sufficient statistic for $\boldsymbol{\theta}$. This completes the proof.

Note: From the proof, the Fisher-Neyman factorization gives us a formula for the conditional probability of $\boldsymbol{x}$ given $\boldsymbol{t}$. In the discrete case we have

$$\mathrm{f}\left(\boldsymbol{x}\,|\,\boldsymbol{t}\right)=\frac{a(\boldsymbol{x})}{\sum_{T(\boldsymbol{x})=\boldsymbol{t}}a(\boldsymbol{x})}$$

An analogous formula holds for continuous random variables (Scharf, pp.82).

Further Examples

The following exercises provide additional examples where the Fisher-Neyman factorization may be used to identify sufficient statistics.

Exercise:

Problem:
Uniform Measurements

Suppose $x_1, \ldots, x_N$ are independent and uniformly distributed on the interval $[\theta_1, \theta_2]$. Find a sufficient statistic for $\boldsymbol{\theta}=(\theta_1\theta_2)^T$.

Note: Express the likelihood $\mathrm{f}_{\boldsymbol{\theta}}\left(\boldsymbol{x}\right)$ in terms of indicator functions.

Exercise:

Problem:
Poisson

Suppose $x_1, \ldots, x_N$ are independent measurements of a Poisson random variable with intensity parameter θ:

$$\forall x, x=0,1,2,\ldots:\left(\mathrm{f}_{\boldsymbol{\theta}}\left(x\right)=\frac{e^{-\theta}\theta^x}{x!}\right)$$

Find a sufficient statistic $\boldsymbol{t}$ for θ.

What is the conditional probability mass function of $\boldsymbol{x}$, given $\boldsymbol{t}$, where $\boldsymbol{x} = (x_1 \ldots x_N)^T$?

Exercise:

Problem:
Normal with Unknown Mean and Variance

Consider $x_1, \ldots, x_N \sim \mathcal{N}\left(\mu, \sigma^2\right)$, IID, where $\theta_1 = \mu$ and $\theta_2 = \sigma^2$ are both unknown. Find a sufficient statistic for $\boldsymbol{\theta} = (\theta_1 \theta_2)^T$.

Note: Use the same trick as in [link].

Gelfand n-widths

Continuing from last time, we have a signal $x \in \mathbb{R}^N$, an $n \times N$ measurement matrix Φ, and $y = \Phi x \in \mathbb{R}^n$ is the information we draw from x.

We consider decoders Δ mapping $\mathbb{R}^n \to \mathbb{R}^N$. We have been discussing whether there exists a decoder with certain properties. So for this discussion (about information preservation), we can just think about optimal decoding.

While the previous result on sparse signal recovery is interesting, it is not very robust. What happens if our signal does not have support k? Can we still obtain meaningful results? The answer is yes. To do so we introduce more general input signal classes $K \subseteq X$ that allows fully supported signals. For example, we will consider the signal class defined by the unit ball

Equation:

$$K = U\left(\ell_p^N\right) = \left\{x : \| x \|_X \leq 1\right\}.$$

Given an encoder/decoder pair (Φ, Δ), the worst case error on a set K for that pair will be given by

Equation:

$$E(K,\Phi,\Delta)_X = \sup_{x \in K} \| x - \Delta(\Phi x) \|_X.$$

Finally, using min-max principles we will define the minimum error over all encoder/decoder pairs for a signal class and for a fixed number of measurements n to be

Equation:

$$E_n(K)_X = \inf_{(\Phi,\Delta):\Phi \text{ is } n \times N} E(K,\Phi,\Delta)_X.$$

This measure $E_n(K)_X$ is the best we could do while measuring distortion on the topology of X, using n linear measurements, and using arbitrary decoding.

We will see that these questions are actually equivalent to a classical “n-width” problem. n-widths have seen a great deal of work over the years by a variety of mathematicians: Kolmogorov, Tikhomirov, Kashin, Gluskin, etc. There are many different flavors of n-widths, but we will study the Gelfand n-width (the least intuitive of the widths).

Definition

Gelfand n-width

Let $K \subseteq X$ be compact. Given n, the Gelfand width (also called the dual width) is given by

Equation:

$$d^n(K)_X := \inf_{Y:\operatorname{codim}(Y)=n} \sup \left\{ \parallel x \parallel_X : x \in K \cap Y \right\}.$$

where by codimension $(Y)=n$ we mean that Y has dimension dim $(X) - n$.

In other words, we are looking for the subspace Y that slices through the set K so that the norms of the projected signals are as small as possible. We can now state the following theorem about n-widths:

Provided that K has the properties (1) $K = -K$ and (2) $K + K = CK$, then

Equation:

$$d^n(K)_X \leq E_n(K)_X \leq C d^n(K)_X$$

where C is the same constant listed in property (2).[footnote]

Clarifying notation: $CK = \{Cx\!:\!x \in K\}$ and $K + K = \{x_1 + x_2\!:\!x_1,x_2 \in K\}$.

We start with the left-hand inequality. We want to take any encoder/decoder pair and use that to construct a Y. So let $\mathit{\Phi}$, $\mathit{\Delta}$ be an encoder/decoder. Then simply let $Y = \mathscr{N}(\mathit{\Phi})$. Now consider an $\eta \in K \cap Y$ and note that $\mathit{\Phi}(\eta) = 0$ since $\eta \in Y$. Let $z = \mathit{\Delta}(0)$ be the decoding of 0 (practically speaking, z should be zero itself, but we avoid that assumption in this proof). Then

Equation:

$$\begin{aligned} \| \eta \|_X &\leq \max\left(\| \eta - z \|_X, \| \eta + z \|_X\right) \\ &= \max\left(\| \eta - \mathit{\Delta}\mathit{\Phi}(\eta) \|_X, \| -\eta - \mathit{\Delta}\mathit{\Phi}(\eta) \|_X\right) \\ &\leq \sup_{x \in K} \| x - \mathit{\Delta}\mathit{\Phi}(x) \|_X \end{aligned}$$

where we first employ the triangle inequality, then the fact that multiplying by -1 does not change the norm, then the fact that $K = -K$. So then

Equation:

$$\sup_{\eta \in K \cap Y} \| \eta \|_X \leq \sup_{x \in K} \| x - \mathit{\Delta}\mathit{\Phi}(x) \|_X.$$

Taking the infimum over all $\mathit{\Phi}$, $\mathit{\Delta}$, it follows that

Equation:

$$d^n(K)_X \leq E_n(K)_X.$$

Since $\mathit{\Phi}$ is $n \times N$, then $\dim(\mathscr{N}(\mathit{\Phi})) \geq N - n$.

Now we prove the right-hand inequality. Assume we have a good Y. Suppose Y has codimension $N - n$. Then $Y^{\perp}$ (the orthogonal complement of Y in $\mathbb{R}^N$) has dimension n. Let $v_1, v_2, \cdots, v_n \in \mathbb{R}^N$ be a basis for $Y^{\perp}$. Let $\mathit{\Phi}$ be the $n \times N$ matrix obtained by stacking the rows $v_1, v_2, \cdots, v_n$. Then $\mathscr{N}(\mathit{\Phi}) = Y$. Define $\mathit{\Delta}(y) =$ any element of $K \cap \mathscr{F}(y)$ if there is one (otherwise let $\mathit{\Delta}(y)$ be anything in $\mathscr{F}(y)$). Now look at the performance $\| x - \mathit{\Delta}\mathit{\Phi}(x) \|_X$ for some $x \in K$. Both x and $\mathit{\Delta}\mathit{\Phi}(x) =: x'$

are elements of K, so $x - x'$ is in $\mathcal{N}(\Phi)$ and in CK. Therefore $\frac{x-x'}{C} \in K \cap \mathcal{N}(\Phi)$. Thus,

Equation:

$$\| \frac{x - x'}{C} \|_X \leq \sup_{z \in Y \cap K} \| z \|_X,$$

and so for any $x \in K$,

Equation:

$$\| x - \Delta\Phi \left(x \right) \|_X \leq C \sup_{z \in Y \cap K} \| z \|_X.$$

Taking the infimum over all Y, we get that $E_n(K)_X \leq Cd^n(K)_X$.

From the proof of this theorem, we see that there is a matching between the matrices Φ and the spaces Y (via the nullspace of Φ).

An important result is that $d^n\left(U\left(\ell_q^N\right)\right)_{\ell_p^N}$ is known for all p, q except $p = 1, q = \infty$. A precise statement of these widths can be found in the book [link]. A particularly important case is

Equation:

$$C_1 \sqrt{\frac{\log(N/n)}{n}} \leq d^n\left(U\left(\ell_1^N\right)\right)_{\ell_2^N} \leq C_2 \sqrt{\frac{\log(N/n)}{n}}$$

for $N > 2n$. This result was first proved by Kashin with a worse logarithm in the upper inequality and later brought to the present form by Gluskin. This result solves several important problems in functional analysis and approximation.

Molecular Structure and Physical Properties

Foundation

We begin with our knowledge of the structure and properties of atoms. We know that atoms have a nuclear structure, meaning that all of the positive charge and virtually all of the mass of the atom are concentrated in a nucleus which is a very small fraction of the volume of the atom. In addition, we know that many of the properties of atoms can be understood by a model in which the electrons in the atom are arranged in “shells” about the nucleus, with each shell farther from the nucleus that the previous. The electrons in outer shells are more weakly attached to the atom than the electrons in the inner shells, and only a limited number of electrons can fit in each shell. Within each shell are subshells, each of which can also hold a limited number of electrons. The electrons in different subshells have different energies and different locations for motion about the nucleus. We also assume a knowledge of the a **Lewis structure** model for chemical bonding based on valence shell electron pair sharing and the octet rule. A covalent chemical bond is formed when the two bonded atoms share a pair of valence shell electrons between them. In general, atoms of Groups IV through VII bond so as to complete an octet of valence shell electrons. We finally assume the **Electron Domain Model** for understanding and predicting molecular geometries. The pairs of valence shell electrons are arranged in bonding and non-bonding domains, and these domains are separated in space to minimize electron-electron repulsions. This electron domain arrangement determines the molecular geometry.

Goals

We should expect that the properties of molecules, and correspondingly the substances which they comprise, should depend on the details of the structure and bonding in these molecules. Now that we have developed an understanding of the relationship between molecular structure and chemical bonding, we analyze physical properties of the molecules and compounds of these molecules to relate to this bonding and structure. Simple examples of physical properties which can be related to molecular properties are the melting and boiling temperatures. These vary dramatically from substance to substance, even for substances which appear similar in molecular formulae, with some melting temperatures in the hundreds or thousands of degrees Celsius and others well below 0°C. We seek to understand these variations by analyzing molecular structures.

To develop this understanding, we will have to apply more details of our understanding of atomic structure and electronic configurations. In our covalent bonding model, we have assumed that atoms “share” electrons to form a bond. However, our knowledge of the properties of atoms reveals that different atoms attract electrons with different strengths, resulting in very strong variations in ionization energies, atomic radii, and electron affinities. We seek to incorporate this information into our understanding of chemical bonding

Observation 1: Compounds of Groups I and II

We begin by analyzing compounds formed from elements from Groups I and II (e.g. sodium and magnesium). These compounds are not currently part of our Lewis structure model. For example, Sodium, with a single valence electron, is unlikely to gain seven additional electrons to complete an octet. Indeed, the common valence of the alkali metals in Group I is 1, not 7, and the common valence of the alkaline earth metals is 2, not 6. Thus, our current model of bonding does not apply to elements in these groups.

To develop an understanding of bonding in these compounds, we focus on the halides of these elements. In [link], we compare physical properties of the chlorides of elements in Groups I and II to the chlorides of the elements of Groups IV, V, and VI, and we see enormous differences. All of the alkali halides and alkaline earth halides are solids at room temperature and have melting points in the hundreds of degrees centigrade. The melting point of $\mathrm{Na\,Cl}$ is 808°C, for example. By contrast, the melting points of the non-metal halides from Periods 2 and 3, such as $C\,\mathrm{Cl}_4$, $P\,\mathrm{Cl}_3$, and $S\,\mathrm{Cl}_2$, are below 0°C, so that these materials are liquids at room temperature. Furthermore, all of these compounds have low boiling points, typically in the range of 50°C to 80°C.

	Melting Point (°C)	**Boiling Point (°C)**
$\mathrm{Li}\,\mathrm{Cl}$	610	1382
$\mathrm{Be}\,\mathrm{Cl}_2$	405	488
$C\,\mathrm{Cl}_4$	-23	77
$N\,\mathrm{Cl}_3$	-40	71
$O\,\mathrm{Cl}_2$	-20	4
$F\,\mathrm{Cl}$	-154	-101
$\mathrm{Na}\,\mathrm{Cl}$	808	1465
$\mathrm{Mg}\,\mathrm{Cl}_2$	714	1418
$\mathrm{Si}\,\mathrm{Cl}_4$	-68	57
$P\,\mathrm{Cl}_3$	-91	74
$S\,\mathrm{Cl}_2$	-122	59
Cl_2	-102	-35
$K\,\mathrm{Cl}$	772	1407
$\mathrm{Ca}\,\mathrm{Cl}_2$	772	>1600

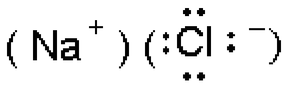
Melting Points and Boiling Points of Chloride Compounds

Second, the non-metal halide liquids are electrical insulators, that is, they do not conduct an electrical current. By contrast, when we melt an alkali halide or alkaline earth halide, the resulting liquid is an excellent electrical conductor. This indicates that these molten compounds consist of ions, whereas the non-metal halides do not.

We must conclude that the bonding of atoms in alkali halides and alkaline earth halides differs significantly from bonding in non-metal halides. We need to extend our valence shell electron model to account for this bonding, and in particular, we must account for the presence of ions in the molten metal halides. Consider the prototypical example of $NaCl$. We have already deduced that Cl atoms react so as to form a complete octet of valence shell electrons. Such an octet could be achieved by covalently sharing the single valence shell electron from a sodium atom. However, such a covalent sharing is clearly inconsistent with the presence of ions in molten sodium chloride. Furthermore, this type of bond would predict that $NaCl$ should have similar properties to other covalent chloride compounds, most of which are liquids at room temperature. By contrast, we might imagine that the chlorine atom completes its octet by taking the valence shell electron from a sodium atom, without covalent sharing. This would account for the presence of Na^+ and Cl^- ions in molten sodium chloride.

In the absence of a covalent sharing of an electron pair, though, what accounts for the stability of sodium chloride as a compound? It is relatively obvious that a negatively charged chloride ion will be attracted electrostatically to a positively charged sodium ion. We must also add to this model, however, the fact that individual molecules of $NaCl$ are not generally observed at temperatures less than 1465°C, the boiling point of sodium chloride. Note that, if solid sodium chloride consists of individual sodium ions in proximity to individual chloride ions, then each positive ion is not simply attracted to a single specific negative ion but rather to all of the negative ions in its near vicinity. Hence, solid sodium chloride cannot be viewed as individual $NaCl$ molecules, but must be viewed rather as a lattice of positive sodium ions interacting with negative chloride ions. This type of “ionic” bonding, which derives from the electrostatic attraction of interlocking lattices of positive and negative ions, accounts for the very high melting and boiling points of the alkali halides.

We can now draw modified Lewis structures to account for ionic bonding, but these are very different from our previous drawings. Sodium chloride can be represented as shown in [link].



This indicates explicitly that the bonding is due to positive-negative ion attraction, and not due to sharing of an electron pair. The only sense in which the Na^{+} ion has obeyed an octet rule is perhaps that, in having emptied its valence shell of electrons, the remaining outer shell of electrons in the ion has the same octet as does a neon atom. We must keep in mind, however, that the positive sodium ion is attracted to many negative chloride ions, and not just the single chloride ion depicted in the Lewis structure.

Observation 2: Molecular Dipole Moments

Our Lewis model of bonding, as currently developed, incorporates two extreme views of the distribution of electrons in a bond. In a covalent bond, we have assumed up to this point that the electron pair is shared perfectly. In complete contrast, in ionic bonding we have assumed that the electrons are not shared at all. Rather, one of the atoms is assumed to entirely extract one or more electrons from the other. We might expect that a more accurate description of the reality of chemical bonds falls in general somewhere between these two extremes. To observe this intermediate behavior, we can examine molecular dipole moments.

An electric dipole is a spatial separation of positive and negative charges. In the simplest case, a positive charge Q and a negative charge $-Q$ separated by a distance R produce a measurable **dipole moment**, μ equal to $Q \times R$. An electric field can interact with an electric dipole and can even orient the dipole in the direction of the field.

We might initially expect that molecules do not in general have dipole moments. Each atom entering into a chemical bond is electrically neutral, with equal numbers of positive and negative charges. Consequently, a molecule formed from neutral atoms must also be electrically neutral. Although electron pairs are shared between bonded nuclei, this does not affect the total number of negative charges. We might from these simple statements that molecules would be unaffected by electric or magnetic fields, each molecule behaving as a single uncharged particle.

This prediction is incorrect, however. To illustrate, a stream of water can be deflected by an electrically charged object near the stream, indicating that individual water molecules exhibit a dipole moment. A water molecule is rather more complicated than a simple separation of a positive and negative charges, however. Recall though that a water molecule has equal total numbers of positive and negative charges, consisting of three positively charged nuclei surrounded by ten electrons. Nevertheless, measurements reveal that water has a dipole moment of $6.17 \times 10^{-30}\,(Cm) = 1.85\,\text{debye}$. (The debye is a unit used to measure dipole moments: $1\,\text{debye} = 3.33 \times 10^{-30}\,(Cm)$.) Water is not unique: the molecules of most substances have dipole moments. A sampling of molecules and their dipole moments is given in [link].

	μ (debye)
H_2O	1.85
HF	1.91
$H\,\text{Cl}$	1.08
$H\,\text{Br}$	0.80
HI	0.42
CO	0.12
CO_2	0
NH_3	1.47
PH_3	0.58

	μ (debye)
$\mathrm{As}\,H_3$	0.20
CH_4	0
$\mathrm{Na\,Cl}$	9.00

Dipole Moments of Specific Molecules

Focusing again on the water molecule, how can we account for the existence of a dipole moment in a neutral molecule? The existence of the dipole moment reveals that a water molecule must have an internal separation of positive partial charge δ and negative partial charge $-\delta$. Thus, it must be true that the electrons in the covalent bond between hydrogen and oxygen are not **equally** shared. Rather, the shared electrons must spend more time in the vicinity of one nucleus than the other. The molecule thus has one region where, on average, there is a net surplus of negative charge and one region where, on average, there is a compensating surplus of positive charge, thus producing a molecular dipole. Additional observations reveal that the oxygen "end" of the molecule holds the partial negative charge. Hence, the covalently shared electrons spend more time near the oxygen atom than near the hydrogen atoms. We conclude that oxygen atoms have a greater ability to attract the shared electrons in the bond than do hydrogen atoms.

We should not be surprised by the fact that individual atoms of different elements have differing abilities to attract electrons to themselves. We have previously seen that different atoms have greatly varying ionization energies, representing great variation in the extent to which atoms cling to their electrons. We have also seen great variation in the electron affinities of atoms, representing variation in the extent to which atoms attract an added electron. We now define the **electronegativity** of an atom as the ability of the atom to attract electrons in a chemical bond. This is different than either ionization energy or electron affinity, because electronegativity is the attraction of electrons **in a chemical bond**, whereas ionization energy and electron affinity refer to removal and attachment of electrons in free atoms.

However, we can expect electronegativity to be correlated with electron affinity and ionization energy. In particular, the electronegativity of an atom arises from a combination of properties of the atom, including the size of the atom, the charge on the nucleus, the number of electrons about the nuclei, and the number of electrons in the valence shell.

Because electronegativity is an abstractly defined property, it cannot be directly measured. In fact, there are many definitions of electronegativity, resulting in many different scales of electronegativities. However, relative electronegativities can be observed indirectly by measuring molecular dipole moments: in general, the greater the dipole moment, the greater the separation of charges must be, and therefore, the less equal the sharing of the bonding electrons must be.

With this in mind, we refer back to the dipoles given in [link]. There are several important trends in these data. Note that each hydrogen halide (HF, $H\text{Cl}$, $H\text{Br}$, and HI) has a significant dipole moment. Moreover, the dipole moments increase as we move **up** the periodic table in the halogen group. We can conclude that fluorine atoms have a greater electronegativity than do chlorine atoms, etc. Note also that HF has a greater dipole moment than H_2O, which is in turn greater than that of NH_3. We can conclude that electronegativity increases as we move **across** the periodic table from left to right in a single period. These trends hold generally in comparisons of the electronegativities of the individual elements. One set of relative electronegativities of atoms in the first three rows of the periodic table is given in [link].

	χ
H	2.1
He	-

	χ
Li	1.0
Be	1.5
B	2.0
C	2.5
N	3.0
O	3.5
F	4.0
Ne	-
Na	0.9
Mg	1.2
Al	1.5
Si	1.8
P	2.1
S	2.5
Cl	3.0
Ar	-
K	0.8

	χ
Ca	1.0

Electronegativities of Selected Atoms

Observation 3: Dipole Moments in Polyatomic Molecules

We might reasonably expect from our analysis to observe a dipole moment in any molecule formed from atoms with different electronegativities. Although this must be the case for a diatomic molecule, this is not necessarily true for a polyatomic molecule, i.e. one with more than two atoms. For example, carbon is more electronegative than hydrogen. However, the simplest molecule formed from carbon and hydrogen (e.g. CH_4) does **not** possess a dipole moment, as we see in [link]. Similarly, oxygen is significantly more electronegative than carbon, yet CO_2 is a non-polar molecule. An analysis of molecular dipole moments in polyatomic molecules requires us to apply our understanding of molecular geometry.

Note that each CO bond is expected to be polar, due to the unequal sharing of the electron pairs between the carbon and the oxygen. Thus, the carbon atom should have a slight positive charge and the oxygen atom a slight negative charge in each CO bond. However, since each oxygen atom should have the same net negative charge, neither end of the molecule would display a greater affinity for an electric field. Moreover, because CO_2 is linear, the dipole in one CO bond is exactly offset by the dipole in the opposite direction due to the other CO bond. As measured by an electric field from a distance, the CO_2 molecule does not appear to have separated positive and negative charges and therefore does not display polarity. Thus, in predicting molecular dipoles we must take into account both differences in electronegativity, which affect bond polarity, and overall molecular geometry, which can produce cancellation of bond polarities.

Using this same argument, we can rationalize the zero molecular dipole moments observed for other molecules, such as methane, ethene and acetylene. In each of these molecules, the individual CH bonds are polar.

However, the symmetry of the molecule produces a cancellation of these bond dipoles overall, and none of these molecules have a molecular dipole moment.

As an example of how a molecular property like the dipole moment can affect the macroscopic property of a substance, we can examine the boiling points of various compounds. The boiling point of a compound is determined by the strength of the forces between molecules of the compound: the stronger the force, the more energy is required to separate the molecules, the higher the temperature required to provide this energy. Therefore, molecules with strong intermolecular forces have high boiling points.

We begin by comparing molecules which are similar in size, such as the hydrides $\mathrm{Si}H_4$, PH_3, and SH_2 from the third period. The boiling points at standard pressure for these molecules are, respectively, -111.8°C, -87.7°C, and -60.7°C. All three compounds are thus gases at room temperature and well below. These molecules have very similar masses and have exactly the same number of electrons. However, the dipole moments of these molecules are very different. The dipole moment of $\mathrm{Si}H_4$, is 0.0D, the dipole moment of PH_3 is 0.58D, and the dipole moment of SH_2 is 0.97D. Note that, for these similar molecules, the higher the dipole moment, the higher the boiling point. Thus, molecules with larger dipole moments generally have stronger intermolecular forces than similar molecules with smaller dipole moments. This is because the positive end of the dipole in one molecule can interact electrostatically with the negative end of the dipole in another molecules, and vice versa.

We note, however, that one cannot generally predict from dipole moment information only the relative boiling points of compounds of very dissimilar molecules

Review and Discussion Questions

Exercise:

Problem:

Compare and contrast the chemical and physical properties of $K\,\mathrm{Cl}$ and $C\,\mathrm{Cl}_4$, and compare and contrast how the chemical bonding model can be used to account for these properties.

Exercise:

Problem: Why is the dipole moment of $\mathrm{Na}\,\mathrm{Cl}$ extremely large?

Exercise:

Problem: Explain why CO has a dipole moment but CO_2 does not.

Exercise:

Problem:

Explain why an atom with a high ionization energy is expected to have a high electronegativity. Explain why an atom with a high electron affinity is expected to have a high electronegativity.

Exercise:

Problem:

Would you predict that a Kr atom has high electronegativity or low electronegativity? Predict the relative electronegativity of Kr and F.

Exercise:

Problem:

Explain why S has a greater electronegativity than P but a smaller electronegativity than O.

Exercise:

Problem:

N atoms have a high electronegativity. However, N atoms have no electron affinity, meaning that N atoms do not attract electrons. Explain how and why these facts are not inconsistent.

Exercise:

Problem:

Explain why compounds formed from elements with large differences in electronegativities are ionic.

Exercise:

Problem:

Explain why ionic compounds have much higher melting points than covalent compounds.

The Environmental Impact of the Manufacturing of Seminconductors (ticket #5622)

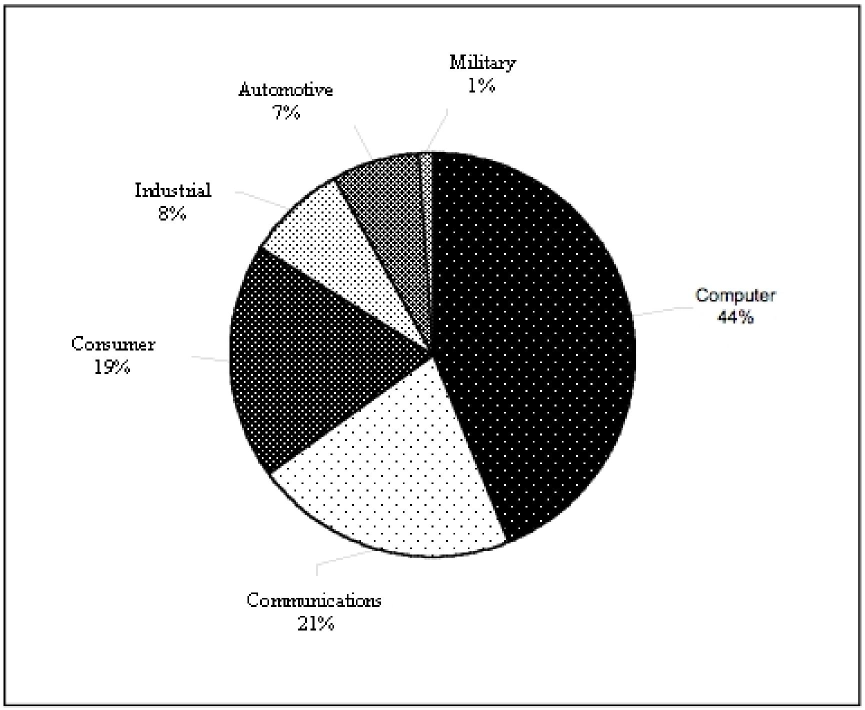
This module gives a brief general overview of semi-conductor manufacturing and some of the components and processes used to produce them that can potentially cause harm to humans or the environment.

Note:"This module was developed as part of a Rice University Class called "Nanotechnology: Content and Context"initially funded by the National Science Foundation under Grant No. EEC-0407237. It was conceived, researched, written and edited by students in the Fall 2005 version of the class, and reviewed by participating professors."

What is a semiconductor?

The semiconductor industry is one of the fastest growing manufacturing sectors in not only the United States but also in the world. According to the American Electronics Association, the domestic sales of electronic components have skyrocketed, jumping from $127 billion to $306 billion over the course of the 1980’s. In the first three quarters of the 2003 fiscal year alone, the export of technology goods from the United States increased by $19 billion [1].

The word “semiconductor” technically refers to any member of a class of solid, crystalline materials that is characterized by an electrical conductivity better than that of insulators (e.g., plastic) but less than that of good conductors (e.g., copper) [2]. Semiconductors are particularly useful as a base material in the manufacturing of computer chips, and the term semiconductor has actually come to be synonymous with the computer chips, themselves. However, semiconductors are not only used in computers. Computers only make up 44% of entire industry consumption (see [link]). Semiconductors are also used for military, automotive, industrial, communications, and other consumer purposes.

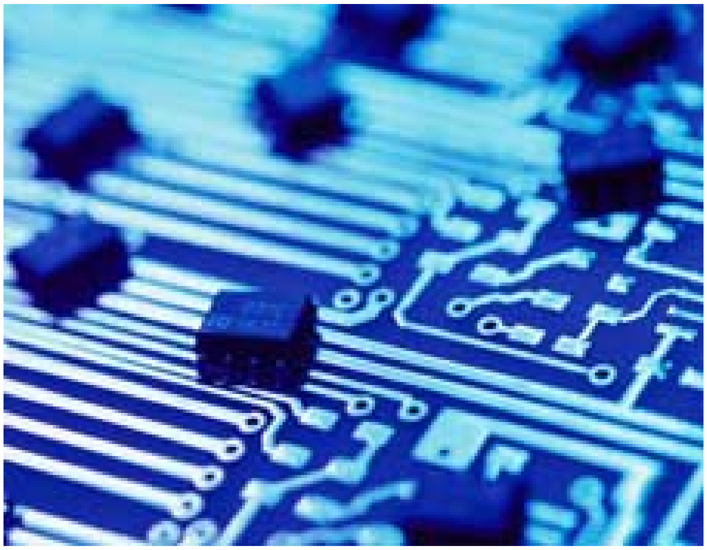


Relative consumption of semiconductors by industry [3].

Semiconductors seem to be anywhere and everywhere throughout our everyday lives, yet it is surprising how little most people know about how they actually work or about the potentially devastating effects their manufacturing can have on the environment and human health.

Why is nanotechnology important to the semiconductor industry?

Much of the study of nanotechnology has been centered on the manufacturing of semiconductors. Though there are a number of highly anticipated applications for nanotechnology in other fields, notably in medicine and in biotechnology, the most tangible results thus far can be argued to have been achieved in the semiconductor industry.



An example of a semiconductor
(photo from PEAK).

For example, Intel recently unveiled its first products based on a generation of 90-nanometer process technology, and its researches and engineers have built and tested prototype transistors all the way down to the 22-nanometer range. Currently, Intel scientists and engineers are working on identifying new materials such as carbon nanotubes and nanowires to replace current transistors, and in particular they hope to develop a “tri-gate” transistor approach that would enable chip designers to build transistors below the 22-nanometer range [4].

With the advent of nanotechnology, these transistors are becoming even faster and more powerful, and in accordance with the law of accelerating returns, the industry has been producing smaller transistors at lower costs with each and every passing year. As these semiconductors become smaller and smaller, they are quickly and surely pushing towards the limits of the nano-realm.

These innovations, however, do not come without their fair share of challenges. Physical issues that are not problematic at the micron scale arise at the nano-scale due to the emergence of quantum effects, and in much the same way that optical microscopy cannot be utilized at the nano-scale, the semiconductor industry is fast approaching a similar diffraction limit. Optical lithography, for instance, a process that uses the properties of light to etch transistors onto wafers of silicon, will soon reach its limit.

At its most basic level, nanotechnology involves pushing individual atoms together one by one. Since approximately 1.7 billion transistors are required for a single chip, this is obviously not a realistic method for mass production. Unless an alternative method for production or a solution to this problem is found, the development and manufacturing of transistors are expected to hit a proverbial brick wall by the year 2015. This is the reason that research in nanotechnology is so important for the world and future of semiconductors.

How are semiconductors manufactured?

Today’s semiconductors are usually composed of silicon, and they are manufactured in a procedure that combines the familiar with the bizarre; some steps that are involved in the process are as everyday as developing a roll of photographic film while others seem as if they would be better suited to take place on a spaceship.

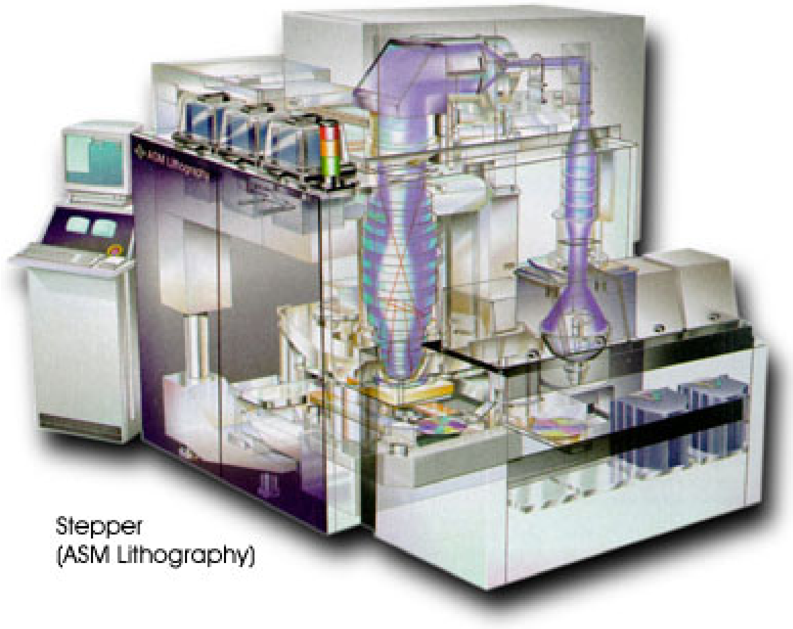
These semiconductors appear to the naked eye as being small and flat, but they are actually three-dimensional “sandwiches” that are ten to twenty layers thick. It can take more than two dozen steps and up to two full months to produce a single one of these silicon sandwiches. Some of the basic and more essential steps involved in the manufacturing process of silicon chips are briefly detailed below.

First, silicon crystals are melted in a vat and purified to 99.9999% purity. The molten silicon is drawn into long, heavy, cylindrical ingots, which are then cut into thin slices called wafersabout the thickness of a business card.

One side of each wafer must be polished absolutely smooth. This process is called chemical-mechanical polishing, and it involves bathing the wafers in special abrasive chemicals. After chemical-mechanical polishing, imperfections cannot be detected on the wafers even with the aid of a laboratory microscope.

After a wafer is polished, layers of material must be stacked on top of the silicon wafer base. Insulating layers are laid down in alternation with conducting layers in a process called deposition. This is often achieved by spraying the chemicals directly onto the surface of the wafer through chemical vapor deposition. Following deposition, the wafer is coated with another layer of chemicals called a photoresist that is sensitive to light.

Next, a machine called a stepper ([link]) is calibrated to project an extremely fine and focused image through a special type of reticle film in a manner similar to that of a simple slide projector. The light that is transmitted through the reticle is projected onto the photoresist layer, which reacts to the light and begins to harden. All of the parts of the wafer exposed to this light harden into a tough crust while the parts in shadow remain soft. This particular step is known by the name of photoelectrochemical etching because it achieves an etching effect, resulting in a chip.



An artist’s illustration of a stepper (image from Solid State Electronics).

Hundreds of copies of the chip are etched onto the wafer until the entire surface has been exposed. Once this process is complete, the entire wafer is submerged into an etching bath, which washes away any parts of the photoresist that remain unexposed along with the insulating chemicals underneath. The hardened areas of the photoresist, however, remain and protect the layers of material underneath them. This process of depositing chemicals, coating with a photoresist, exposure to light over a film mask, and etching and washing away is repeated more than a dozen times. The result is an elaborate, three-dimensional construction of interlocking silicon wires.

This product is then coated with another insulating layer and is plated with a thin layer of metal, usually either aluminum or copper. Yet another photoresist is laid down on top of this metal plating, and after the wafer is exposed in a stepper, the process repeats with another layer of metal. After this step has been repeated several more times, a final wash step is performed, and a finished semiconductor product rolls off the assembly line, at last.

What is a clean room?

A typical semiconductor fabrication facility, or “fab” in industry jargon, looks like a normal two- or three-story office building from the outside, and most of the interior space is devoted to one or more “clean rooms,” in which the semiconductors are actually made. A clean room is designed with a fanatical attention to detail aimed towards keeping the room immaculate and dust-free ([link]).

An industry clean room at AP Tech (photo from Napa Gateway).

Most if not all surfaces inside these clean rooms are composed of stainless steel, and these surfaces are sloped whenever possible or perforated by grating to avoid giving dust a place to settle. The air is filtered through both the ceiling and the floor to remove particles that are down to 1/100 the width of a human hair. Lighting is characteristically bright and slightly yellowish to prevent mildew from forming behind equipment or in recessed corners, and even the workers in a clean room must be absolutely spotless.

Workers in these rooms must be covered from head to toe in “bunny suits” that completely seal the body in a bulky suit, helmet, battery pack, gloves, and boots. Once sealed in these suits, the workers often look more like space explorers in a science fiction movie than computer chip employees, but in order to even enter the stainless steel locker room to suit up to begin with, they must first pass through a series of air lock doors, stand under a number of “air showers” that actually blow dust off of clothing, and walk across a sticky floor matting that removes grime from the bottom of shoes.

Semiconductor-manufacturing companies often portray their fabrication facilities as being clean, environmentally friendly, and conspicuously free of the black, billowing smokestacks that have come to be associated with the plants and factories of other major industries. These facilities produce no visible pollution and certainly do not appear to pose any health or environmental risks.

In truth, the term “clean room,” itself is more than just a bit of an understatement. Industry executives often boast that their clean rooms are from 1,000 times to 10,000 times cleaner and more sanitary than any hospital operating room.

What are the health risks involved in the semiconductor industry?

The use of sterile techniques and the fastidious attention devoted to cleanliness in the semiconductor industry may perpetuate the illusion that the manufacturing of semiconductors is a safe and sterile process. However, as a rapidly growing body of evidence continues to suggest, hardly anything could be further from the truth ([link]). The question of worker safety and chemical contamination at chip-making plants has received an increasing amount of attention over the course of the past decade.

Chemicals used in the manufacturing of semiconductors are known to have toxic effects (image from FARSHA).

The devices being built at semiconductor fabrication facilities are super-sensitive to environmental contaminants. Because each chip takes dozens of trained personnel several weeks to complete, an enormous amount of time and effort is expended to produce a single wafer. The industry may pride itself on its perfectly immaculate laboratories and its bunny-suited workers, but it should be noted that the bunny suits are not designed to protect their wearers from hazardous materials but rather to protect the actual semiconductor products from coming into contact with dirt, hair, flakes of skin, and other contaminants that can be shed from human bodies. They protect the silicon wafers from the people, not the people from the chemicals.

Lee Neal, the head of safety, health, and environmental affairs for the Semiconductor Industry Association, has been quoted as saying, “This is an environment that is cleaner than an operating room at a hospital.” However, this boast is currently being challenged by industry workers, government scientists, and occupational-health experts across the country and worldwide.

Industrial hygiene has always been an issue in the semiconductor industry. Many of the chemicals involved in the manufacturing process of semiconductors are known human carcinogens or pose some other serious health risk if not contained properly. [link] lists ten of the hazardous chemicals most commonly used in manufacturing semiconductors along with their known effects on human health.

Chemical name	**Role in manufacturing process**	**Health problems linked to exposure**
Acetone	Chemical-mechanical polishing of silicon wafers	Nose, throat, lung, and eye irritation, damage to the skin, confusion, unconsciousness, possible coma
Arsenic	Increases conductivity of semiconductor material	Nausea, delirium, vomiting, dyspepsia, diarrhea, decrease in erythrocyte and leukocyte production, abnormal heart rhythm, blood vessel damage, extensive tissue damage to nerves, stomach, intestine, and skin, known human carcinogen for lung cancer
Arsine	Chemical vapor deposition	Headache, malaise, weakness, vertigo, dyspnea, nausea, abdominal and back pain, jaundice, peripheral neuropathy, anemia
Benzene	Photoelectrochemical	Damage to bone

	etching	marrow, anemia, excessive bleeding, immune system effects, increased chance of infection, reproductive effects, known human carcinogen for leukemia
Cadmium	Creates “holes” in silicon lattice to create effect of positive charge	Damage to lungs, renal dysfunction, immediate hepatic injury, bone defects, hypertension, reproductive toxicity, teratogenicity, known human carcinogen for lung and prostate cancer
Hydrochloric acid	Photoelectrochemical etching	Highly corrosive, severe eye and skin burns, conjunctivitis, dermatitis, respiratory irritation
Lead	Electroplated soldering	Damage to renal, reproductive, and immune systems, spontaneous abortion, premature birth, low birth

		weight, learning deficits in children, anemia, memory effects, dementia, decreased reaction time, decreased mental ability
Methyl chloroform	Washing	Headache, central nervous system depression, poor equilibrium, eye, nose, throat, and skin irritation, cardiac arrhythmia
Toluene	Chemical vapor deposition	Weakness, confusion, memory loss, nausea, permanent damage to brain, speech, vision, and hearing problems, loss of muscle control, poor balance, neurological problems and retardation of growth in children, suspected human carcinogen for lung and liver cancer
Trichloroethylene	Washing	Irritation of skin, eyes, and respiratory tract, dizziness,

		drowsiness, speech and hearing impairment, kidney disease, blood disorders, stroke, diabetes, suspected human carcinogen for renal cancer

Chemicals of concern in the semiconductor industry [5].

Several semiconductor manufacturers including National Semiconductor and IBM have been cited in the past for holes in their safety procedures and have been ordered to tighten their handling of carcinogenic and toxic materials.

In 1996, 117 former employees of IBM and the families of 11 workers who had died of cancer filed suit against the chemical manufacturers Eastman Kodak Company, Union Carbide Corporation, J. T. Baker, and KTI Chemicals, claiming that they had suffered adverse health effects as a result of exposure to hazardous chemicals on the job in the semiconductor industry [5]. The lawsuit was filed in New York, which prevented the employees from suing IBM directly. A separate group of former IBM workers who had developed cancer filed suit against the company in California, alleging that they had been exposed to unhealthy doses of carcinogenic chemicals over the past three decades. Witnesses who testified in depositions in the New York state court in Westchester County described how monitors that were supposed to warn workers of toxic leaks often did not function because of corrosion from acids and water. They also alleged that supervisors sometimes shut down monitors to maintain production rates. When they lodged complaints with senior officials in the company, they claim to have been told not to “make waves” [6]. Meanwhile, 70 female workers in Scotland sued National Semiconductor Corporation, another U.S.-based company, claiming that they, too, were exposed to carcinogens on the job.

These lawsuits and the resulting publicity prompted a groundbreaking study by the Health and Safety Executive, which commissioned a committee to investigate these allegations [7]. The committee found that there were indeed unusually high levels of breast and other kinds of cancer among workers at National Semiconductor’s fabrication facility in Greenock, Scotland. The committee concluded that the company had failed to ensure that the local exhaust ventilation systems adequately controlled the potential exposure of employees to hydrofluoric acid and sulphuric acid fumes and to arsenic dust. These findings proved to be extremely embarrassing for the company and for the industry. According to an official statement released by Ira Leighton, acting regional administrator of the New England branch of the U.S. Environmental Protection Agency, "National Semiconductor is a big business that uses a large amount of harmful chemicals and other materials. Our hazardous waste regulations were created to properly monitor dangerous chemicals and prevent spills. In order for it to work, it is important businesses to comply with all of the regulations. When companies fail to do this they are potentially putting people – their employees and neighbors – at risk [8]. "

Moreover, a study of fifteen semiconductor manufacturers published in the December 1995 issue of the American Journal of Independent Medicine showed that women working in the so-called clean rooms of the semiconductor fabs suffered from a 14% miscarriage rate.

Protesters at a rally staged against IBM (photo from San Francisco Independent Media Center).

The main problem in prosecution is that the industry does not have a single overarching and definitive process for manufacturing, and it is difficult to pinpoint one particular compound as causing a certain health problem because some plants use as many as 300 chemicals. Also, many of the manufacturing processes take place in closed systems, so exposure to harmful substances is often difficult to detect unless monitored on a daily basis.

Executives and spokespeople for the semiconductor industry maintain that any chip workers' cancers and other medical problems are more likely due to factors unrelated to the job, such as family history, drinking, smoking, or eating habits. They also say that over the years, as awareness of chemical hazards has grown, they have made efforts to phase out toxic chemicals and to lower exposure to others. They insist that they use state-of-the-art process equipment and chemical transfer systems that limit or prevent physical exposure to chemicals and point out that the substances used in the semiconductor industry are used in other industries without a major health or safety problem.

What environmental risks are involved?

In theory, attention to cleanliness is in the manufacturer's best interest not only from a health perspective but also from an economic. Many chemicals used in the production process are not expensive in and of themselves; however, the cost of maintaining these materials in an ultra-clean state can be quite high. This encourages the close monitoring of usage, the minimization of consumption, and the development of recycling and reprocessing techniques. Also, the rising costs of chemical disposal are prompting companies to conduct research into alternatives that use more environmentally friendly methods and materials. Individual companies and worldwide trade associations were active in reducing the use and emission of greenhouse gases during the 1990’s, and the industry as a whole has substantially reduced emissions over the last twenty years.

Nonetheless, there has been a history of environmental problems linked to the industry in Silicon Valley and other technology centers. To begin with, a tremendous amount of raw materials is invested in the manufacturing of semiconductors every year.

Moreover, a typical facility producing semiconductors on six-inch wafers reportedly uses not only 240,000 kilowatt hours of electricity but also over 2 million gallons of water every day [9]. Newer facilities that produce eight-inch and twelve-inch wafers consume even more, with some estimates going as high as five million gallons of water daily. While recycling and reusing of water does occur, extensive chemical treatment is required for remediation, and in dry or desert areas such as Albuquerque, New Mexico, home to plants for Motorola, Philips Semiconductor, Allied Signal and Signetics, Intel, and other high-tech firms, the high consumption of water necessary for the manufacturing of semiconductors can pose an especially significant drain on an already scarce natural resource [10]. The existence of economic mainstays including the mining industry and the established presences of Sandia National Laboratories and the Los Alamos National Laboratory make New Mexico an attractive location for high-tech tenants. However, the opening of fabrication facilities in the state leaves its farmers and ranchers in constant competition with the corporations for rights to water consumption. On average, the manufacturing of just 1/8-inch of a silicon wafer requires about 3,787 gallons of wastewater, not to mention 27 pounds of chemicals and 29 cubic feet of hazardous gases [11].

A community near Sutter Creek, California that has been designated as an EPA Superfund site as a result of arsenic contamination (photo from Alexander, Hawes, & Audet).

Contamination has also been an issue in areas surrounding fabrication plants. Drinking water was found to be contaminated with trichloroethane and Freon, toxins commonly used in the semiconductor industry, in San Jose, California in 1981 [12]. These toxins were later suspected to be the cause of birth defects of many children in the area. The culprits were Fairchild Semiconductor and IBM. The companies' underground storage tanks were found to have leaked tens of thousands of gallons of the toxic solvents into the ground. There are a number of semiconductor-related EPA cleanup sites in Silicon Valley, and there have been concerns raised about the cumulative air and groundwater pollution in Silicon Valley, as well.

Another area of concern is the eventual fate of discarded electronic systems such as computers, pagers, mobile phones, and televisions that contain semiconductor devices. Personal computers in particular are especially problematic because they become obsolete fairly rapidly and lose almost all of their market value within five or ten years after their date of manufacture. Tens of millions of PC's are sold in the United States each year, and they pose an environmental risk not only through their sheer bulk in city dumps and landfills but also because their semiconducting devices often contain significant amounts of heavy metals, including lead and other potentially hazardous substances.

Why don’t we hear more about this on the news?

Across the United States, approximately 60% of the manufacturing facilities for semiconductor devices are located in six states. These states listed in descending order are California, Texas, Massachusetts, New York, Illinois, and Pennsylvania. The industry appears to be concentrated in these particular locations in part because they are near the primary users, transportation routes, and experts in the field, but people of all ages in all fifty states are impacted by semiconductor technology. Consumerism of semiconductor products is only expected to increase in coming years. Apple, for instance, expects to have sold 23.6 million iPods, devices that rely on semiconductor technology, by the year 2006.

If semiconductors are so ubiquitous in our day-to-day lives, why is there so little awareness about the serious environmental and health risks that are involved in their manufacturing process? Part of the problem is that little is known about the long-term health or environmental consequences of exposure to the chemicals that are used in the process. Because the semiconductor industry is still relatively new, not many studies have been conducted on this topic, and existing data is often inconclusive. This being said, some scientists predict that the cancer rate in the silicon chip industry will rise significantly in the future because cancer can take as long as 20-25 years to manifest itself in populations of exposed workers.

The EPA does have regulations in effect that are aimed toward the purpose of controlling the levels of contaminants released and minimizing human and environmental exposure to them. However, current regulations do not mandate that American companies report on offshore manufacturing. Therefore, even as media coverage and general awareness increase, companies can simply outsource more and more of their fabrication facilities to, for example, Southeast Asia. Some companies, in fact, have begun to do so, and there have even already been studies conducted on the health issues of workers in the electronics and semiconductor industries of Singapore and Malaysia [13].

Thus, changes in how and where semiconductor firms manufacture chips currently outstrip the present ability of the United States government and media institutions to track and monitor their potential threats to humans and the environment. If this situation is to change for the better in the near future, it is clear that radical reforms will need to take place on a number of different levels. However, the who, what, when, where, and why, so to speak, of that reform remains to be addressed.

Discussion questions

- How many electronics products do you use on a day-to-day basis? How many of these products contain semiconductors?
- Who do you think is ultimately responsible for initiating reform? The government? The corporation? The consumer?
- Do you think that the health and environmental incidents related to semiconductor manufacturing will remain isolated incidents? Or do you think that these incidents will become epidemic in the future?
- Do you think that nanotechnology will help the problem or make the problem worse?

Endnotes

1. M. Kazmierczak and J. James. Industry Data & Publications: U.S. High-Tech Exports, 2000-2004. 16 Nov. 2004. American Electronics Association. 17 Oct. 2005<http://www.aeanet.org/Publications/idjl_ushightechexports1204.asp>.
2. J. Turley, The Essential Guide to Semiconductors. Upper Saddle River, New Jersey: Prentice Hall Professional Technical Reference, 2003.
3. J. Turley, The Essential Guide to Semiconductors. Upper Saddle River, New Jersey: Prentice Hall Professional Technical Reference, 2003. From Prentice Hall
4. IBM Research Nanotechnology Homepage. IBM. 16 Oct. 2005 <http://domino.research.ibm.com/comm/research.nsf/pages/r.nanotech.

html>.

5. R. Chepesiuk, “Where the Chips Fall: Environmental Health in the Semiconductor Industry.” Environmental Health Perspectives 107 (1999): 452-457.
6. Richards, “Industry Challenge: Computer-Chip Plants Aren’t as Safe And Clean As Billed, Some Say – Women at Scottish Factory Tell of Spills and Fumes, Face Host of Medical Ills – Firms Won’t Help Do a Study.” Wall Street Journal 5 Oct. 1998, eastern ed.: A1.
7. A. Heavens, Chip Plants Take Heat For Toxics. 14 Jan. 2003. Wired News. 13 Oct. 2005, <http://www.wired.com/news/technology/0,1282,57191,00.html>
8. M. Merchant, Maine Semiconductor Plant Fined For Hazardous Waste Violations. Boston: U.S. Environmental Protection Agency, Press Office, 2001.
9. P. Dunn, Cleanliness Outside, Some Issues Outside. 2 Oct. 2000. The Foundation for American Communications. 13 Oct. 2005 <http://www.facsnet.org/tools/sci_tech/tech/community/environ2.php3>
10. J. Mazurek, Making Microchips: Policy, Globalization, and Economic Restructuring in the Semiconductor Industry. Cambridge, Massachusetts: MIT Press, 1999.
11. C. Hayhurst, “Toxic Technology: Electronics and the Silicon Valley.” E: the Environmental Magazine May-Jun. 1997: 4.
12. B. Pimentel, “The Valley’s Toxic History.” San Francisco Chronicle 30 Jan. 2004, final ed.: B1.
13. V. Lin, Health, Women’s Work, and Industrialization: Semiconductor Workers in Singapore and Malaysia. New York: Garland Publishing, Inc., 1991.

CNXML 0.5 Stress Test

This brief paragraph was put in before a section tag to ensure that it was styled correctly.

Equation:

This equation has been thrown in as a child of content for styling purposes.

$$\cos(2\pi ft) = \frac{e^{i2\pi ft} + e^{-(i2\pi ft)}}{2}$$

General Example with Equation

The complex exponential defines the notion of frequency: It is the **only** signal that contains only one frequency component. The sinusoid consists of two frequency components: one at the frequency $+f_0$ and the other at $-f_0$.

Equation:

This is Euler's relation relating things like e's and cosines

$$\cos(2\pi ft) = \frac{e^{i2\pi ft} + e^{-(i2\pi ft)}}{2}$$

Equation:

Euler's relation

$$\sin(2\pi ft) = \frac{e^{i2\pi ft} - e^{-(i2\pi ft)}}{2i}$$

Equation:

Another Euler's relation

$$e^{i2\pi ft} = \cos(2\pi ft) + i\sin(2\pi ft)$$

The complex exponential signal can thus be written in terms of its real and imaginary parts using Euler's relation. Thus, sinusoidal signals can be expressed as either the real or the imaginary part of a complex exponential signal (the choice depends on whether cosine or sine phase is needed) or as the sum of two complex exponentials. These two decompositions are mathematically equivalent to each other.

Section: ICSA

The Rice Sailing Team is part of a sailing coalition called ICSA which stands for Inter-Collegiate Sailing Association. The ICSA is the governing body for collegete sailing programs in the US. Volunteers of ICSA develop the racing rules. They also organize the national regattas. To be a member of ICSA, a college must pay money to a sub-district.

SEISA

Rice is a member of the SEISA district of ICSA. SEISA stands for South-Eastern Inter-Collegiate Sailing Association. SEISA is the organization that the South-East colleges compete in. The schools gather to compete for all normal regattas. One or two schools from SEISA will qualify for the National competitions hosted by ICSA.

RUST: Rice University Sailing Team

RUST is Rice's competitive sailing team. Rice is a member of SEISA. This district is made up of several schools, including UT, A&M, A&M Galveston, UNO, Tulane, Kansas, and UWF. The Rice Sailing Club is open to all members of the student community, including graduate students. Beginner and expert sailors alike are encouraged to practice with the team. RUST is a small and easy-going club. There are no attendance requirements, and everyone who wants to sail can sail. Members pay yearly dues to help cover expenses, but every member not only gets to sail, but gets a sailing club T-shirt too!

For the benefit of Mr. Kite, there will be a show tonight on trampoline. The Hendersons will all be there, late Pablo Fanques Fair - what a scene. Over men and horses, hoops and garters, lastly through a hogshead of real fire! In this way Mr. K. will challenge the world! The celebrated Mr. K. performs his feat on Saturday at Bishopsgate. The Hendersons will dance and sing as Mr. Kite flys through the ring - don't be late. Messrs. K and H. assure the public their production will be second to none. And course Henry The Horse dances the waltz!

Equation:

This equation has been thrown in as a child of section for styling purposes

$$\cos(2\pi ft) = \frac{e^{i2\pi ft} + e^{-(i2\pi ft)}}{2}$$

Section Three Point Two

Section 3.2

Section Three Point Three

Section 3.3

four levels down

five levels down

six levels down

seven levels down

eight levels down

nine levels down

how low can you go?

as low as you want

CNXML is limitless in it's powers

You get the picture!

List

Hello and welcome to my list of things to do. The following list is **within** a paragraph, whereas the subsequent ordered list is **preceded** by a paragraph.

- This is an item. The length of this item has been extended for quality assurance of its style upon wrapping, assuming it extends for more than the width of the screen, which is why it may appear that I am dragging this sentence out for a long time..
- This is another item with a list nested inside of it.
    - The next subitem contains a named equation.
    - **Equation:**

      Euler's relation

$$\sin(2\pi ft) = \frac{e^{i2\pi ft} - e^{-(i2\pi ft)}}{2i}$$

This is not a new paragraph. This is also not a new paragraph.

This is a new paragraph.

Example Enumerated List

1. This is an enumerated list item not located within paragraph tags.
2. This is another item with a list nested inside of it.
    - This is a subitem. But the subitem is not enumerated.
    - This is another subitem.
3. Yet another item

A list of notes, definitions, and rules

- **Note:** Sulfuric and Hydrochloric Acid are **very dangerous**. Please handle with caution.

- **Note:** Sulfuric and Hydrochloric Acid are **very dangerous**. Please handle with caution.

- note
    A short letter.
    A different meaning for the term.
    Equation:

$$\sqrt{t} = \sum_{n=-\infty}^{\infty} (-1)^n A \, \mathrm{p}_{T/2} \left(t - n\frac{T}{2} \right)$$

Example:
Bob, Your son called. -Mary

Example:
A first example.

Example:
A second example.

- Pythagorean Theorem

  On a right triangle, the sum of the squares of the sides equals the square of the hypotenus.
  proof name

  Your favorite proof goes here.

Example:
This example is a child of the proof. The next one is not.

Example:
Take a right triangle whose sides are of length 3, 4, and 5. In this case the sum of the square of the two shorter sides is 9+16=25. The square of the hypotenus is 25. So the theorem holds.

This paragraph contains an inline list, i.e. a list with `type="inline`: This item has no name**Name**This is item has a name. Next is an inline list with a name**List name****First item name**First item text2nd item text blah blah blah. This paragraph also contains a named-item list, i.e. a list with `type="named-item"`:

name of list

- This item has no name, even though it's part of a named-item list fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff
- **namenamename**This item as a longish name fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff
- **n**This item has a short name fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff fluff

CNXN and Link

If you want to learn more about things in general, you can look at other pages in CNX world. To learn more about CNXML, go to the CNXML 0.5 spec. Then you can write things like `<cnxn>` and `<link>` in CNXML. You will also be able to make CNXN's between things you wrote and things other people have written. For example, to learn more about Rice sailing, you can look right here at the sailing section in this module. Or you could link to something even less relevant, say digital communication receivers. To go to a topic which is a bit more relavant to writing `<cnxn> ` tags, you could visit the module about using emacs. If you have more questions about Connexions basics, you could look at the FAQ . So if you want to learn more about the world in which you live, Connexions is the place to be.

empty cnxn testing

- To another module: [link]
- To a subfigure: [link]
- To a figure: [link]
- To a paragraph: [link]
- To an equation: [link]
- To a note with no type attribute: [link]
- To a note with type attribute equal to null string: [link]

- To a note with a type attribute other than "note": [link]
- To a footnote: [link]
- To a rule: [link]
- To an emphasis: [link]
- To a definition: [link]
- To an exercise outside of an example: [link]
- To a problem outside an example: [link]
- To a solution outside an example: [link]
- To a solution (one of many) outside an example: [link]
- To an exercise inside an example: [link]
- To a problem inside an example: [link]
- To a solution inside an example: [link]
- To a solution (one of many) inside an example: [link]
- To a glossary: [link]
- To a cnxn: [link]
- To a table caption: [link]
- To a figure caption: [link]
- To an example with no numbering [link]
- To an example after an example with no numbering [link]

Example

Example:
As an example of signal complexity, we can express the pulse $\mathrm{p}_{\Delta}\ (t)$ as a sum of delayed unit steps.

Equation:

$$\mathrm{p}_{\Delta}\ (t) = u(t) - u(t - \Delta)$$

Thus, the pulse is a more complex signal than the step. Be that as it may, the pulse is very useful to us.

Example:

CNXML can be a pain

This sentence, although it is very **unimportant** and **irrelevant**, contains emphasis within the `<example>` tag. For an example of `<cite>` within `<example>`, see *Things That Are a Pain to Style in CNXML* by Max Starkenburg. For an example of `<emphasis>` within `<cite>` within `<example>`, please see *Things That Are a* ***Pain in the Butt*** *to Style in CNXML*, also by Max Starkenburg. This sentence contains **`emphasized codeline`** within an example. This sentence contains a cnxn within an example. This sentence contains a link. This sentence contains a **term**. This sentence contains a **`codeline term`**.

This paragraph has just been thrown in to ensure that it is styled correctly. It has nothing to say, but is long anwyay, just to make sure that it continues to the next line.

Example:

```
          <html>
          <head>
          <title>Codeblock within the example
tag</title>
          </head>

          <body bgcolor="#ffffff" link="#990000">
```

Example:

Example List in Example

1. This is an enumerated list item.
2. This is another item with a list inside of it.

    - This is a bulleted list item

    - This is another bulleted list item

Example:

This figure is inside of the `<example>` tag.

Example:

Transform	Time Domain	Frequency Domain	Convolution
Continuous-Time Fourier Series	$L^2([0,T))$	$l^2(\mathbb{Z})$	Continuous-Time Circular

Transform	Time Domain	Frequency Domain	Convolution
Continuous-Time Fourier Transform	$L^2(\mathbb{R})$	$L^2(\mathbb{R})$	Continuous-Time Linear
Discrete-Time Fourier Transform	$l^2(\mathbb{Z})$	$L^2([0, 2\pi))$	Discrete-Time Linear
Discrete Fourier Transform	$l^2([0, N-1])$	$l^2([0, N-1])$	Discrete-Time Circular

Table of Fourier Representations

Example:

note

A short letter.

Example:

Bob, Your son called. -Mary

Pythagorean Theorem

On a right triangle, the sum of the squares of the sides equals the square of the hypotenus.

proof name

Your favorite proof goes here.

Example:
Take a right triangle whose sides are of length 3, 4, and 5. In this case the sum of the square of the two shorter sides is 9+16=25. The square of the hypotenus is 25. So the theorem holds.

Note: Sulfuric and Hydrochloric Acid are **very dangerous**. Please hande with caution.

Note: Sulfuric and Hydrochloric Acid are **very dangerous**. Please hande with caution.

23h8902
    aj2390tj

Of thumb

jaf3j9w0gea ja jdlfasdnfl

Exercise

Exercise:

Problem:

Express a square wave having period T and amplitude A as a superposition of delayed and amplitude-scaled pulses.

This is an extra paragraph thrown in for styling purposes.

Solution:

This is a paragraph that begins a solution. The equation is nested in the paragraph.

Equation:

$$\sqrt{t} = \sum_{n=-\infty}^{\infty} (-1)^n A \, \mathrm{p}_{T/2} \left(t - n\frac{T}{2} \right)$$

This is an extra paragraph thrown in for styling purposes.

Example:

Exercise:

Problem: Problem Text Here

Solution:

Solution Text Here

Example:

Exercise:

Problem:

Exercise/problem nested inside example nested inside example.

Solution:

Solution nested inside example nested inside example.

section for testing purposes only

Exercise:

Problem: Problem Text Here

Solution:

Solution Text Here

Exercise:

Problem: Problem Text Here

Solution:

Solution 1 Text Here

Solution:

Solution 2 Text Here

Exercise:

Problem:
Equation:

This equation has been thrown in as a child of problem for styling purposes

$$\cos(2\pi ft) = \frac{e^{i2\pi ft} + e^{-(i2\pi ft)}}{2}$$

Solution:

The solution is to use Mozilla, and not IE, nor Netscape. The latter two browsers have serious CSS bugs. For an equation inside a solution, see the above exercise.

Solution:

Note the ability to have more than one solution per exercise.

Exercise:

Problem: This exercise contains a problem but no solution

Example:
Exercise:

Problem: Another exercise in an an example following the other ones

Solution:

Solution text

Codeline and Codeblock

Try running this in MATLAB.

```
                                   for n=1:1000
        y(n) = sum(a.*y(n-1:-1:n-p)) +
sum(b.*x(n:-1:n-q));
        end
```

The `for` command repeats its contents while incrementing a counter.

`<!-- some codeline and codeblock are emphasized more than others -->`

`<!-- codeline can be linked and cnxn'ed too -->`

```
        codeblock
        can be
        linked and
        cnxn'ed
        as well
```

Definition, Term, Meaning, and Example

note
- A short letter.
- A different meaning for the term.

Example:
Bob, Your son called. -Mary

Example:

A first example.

Example:
A second example.

I can also define a new **term** in the middle of a sentence. A term is a new word that is being defined for the first time.

The following definition will be the child of a `<para>` tag to ensure that it is styled well.

`#!/usr/local/bin/perl/`
: The first line of every Perl program

Example:
This program will produce a 2-column listing of a file.

```
#!/usr/local/bin/perl/
while( $1 = >STDIN> ) {
chop( $1 );
if( $1 =~ s/^Desig:\s+// ) {
$desig = $1;
next;
}
if( $1 =~ s/^Area:\s+// ) {
$area = $1;
printf( "%-12s %\n", $desig, $area );
}
}
```

This sentence contains a term which is **half regular** `half codeline`.

Feast your eyes on this: `#!/usr/local/bin/perl/` is a term smack dab in the middle of a sentence! Amazing! This example, `#!/usr/local/bin/perl/` , is a codeline term with emphasis on `/local/`. The next example is a `codeline term with a link and cnxn`.

equation

Equation:

This equation has been thrown in as a child of meaning for styling purposes

$$\cos(2\pi ft) = \frac{e^{i2\pi ft} + e^{-(i2\pi ft)}}{2}$$

Note

Chemistry lab was always a lot of fun. We played with a lot of chemicals such as Sulfuric and Hydrochloric acid.

Note: Sulfuric and Hydrochloric Acid are **very dangerous**. Please handle with caution.

Chemistry lab was always a lot of fun. We played with a lot of chemicals such as Sulfuri and Hydrochloric acid.

Note: This note does **not** have a string in its type attribute. Please handle with caution.

Note: This note does **not** have a type attribute. Please handle with caution.

Chemistry lab was always a lot of fun. We played with a lot of chemicals such as Sulfuric and Hydrocholic acid. [footnote] Adding another footnote here. [footnote]
Sulfuric and Hydrochloric Acid are **very dangerous**. Please handle with caution.
This note was added to test styling.

Chemistry lab was always a lot of fun. We played with a lot of chemicals such as Sulfuric and Hydrocholic acid. [footnote]
Sulfuric and Hydrochloric Acid are **very dangerous**. Please handle with caution. Also handle this note with caution, because it contains the attribute `type="FOOTNOTE"`, not `type="footnote"`, and will thus be styled differently.

Rule

Pythagorean Theorem

On a right triangle, the sum of the squares of the sides equals the square of the hypotenus.
This proof was thrown into to have a child of equation for styling purposes
Equation:

$$\cos(2\pi ft) = \frac{e^{i2\pi ft} + e^{-(i2\pi ft)}}{2}$$

Example:
Take a right triangle whose sides are of length 3, 4, and 5. In this case the sum of the square of the two shorter sides is 9+16=25. The square of the hypotenus is 25. So the theorem holds.

Property

If $s(t)$ is real, $c_{-k} = \overset{*}{c_k}$

This result follows from the integral that calculates the c_k from the signal. Furthermore, this result means that $\operatorname{Re}(c_k) = \operatorname{Re}(c_{-k})$: The real part of the Fourier coefficients for real-valued signals is even. Similarly, $\operatorname{Im}(c_k) = -\operatorname{Im}(c_{-k})$: The imaginary parts of the Fourier coefficients have odd symmetry. Consequently, if you are given the Fourier coefficients for positive indices and zero and are told the signal is real-valued, you can find the negative-indexed coefficients, hence the entire spectrum. This kind of symmetry, $c_{-k} = \overset{*}{c_k}$, is known as **conjugate symmetry**. We can phrase the property concisely by saying:

Property

Real-valued periodic signals have a conjugate-symmetric spectrum.

The Fourier series obeys:

Parseval's Theorem

Power calculated in the time domain equals the power calculated in the frequency domain.

Equation:

$$\frac{1}{T}\int_0^T s^2(t)\,\mathrm{d}\,t = \sum_{k=-\infty}^{\infty} (|c_k|)^2$$

This result is a (simpler) re-expression of how to calculate a signal's power than with the real-valued Fourier series expression for power .

Cite

Useful information about the color blue can be found in *The Big Book of the Color Blue*.

Emphatic information about emphasizing can be found in *The **Gigantic** Book of Emphasis*.

Figures

Here is an example of a figure with a large table inside it. For exhaustive examples of figures, subfigures, and captions in many shapes, sizes, and combinations, see [link]. Currently, figures cannot be placed in paragraph tags.
Ascii table

Income	12,293	12,398	14,298
Expenditure	293	398	298
Surplus	12000	12000	14000

Budget - Comparison by year

And here we can add more information about the table in a caption!

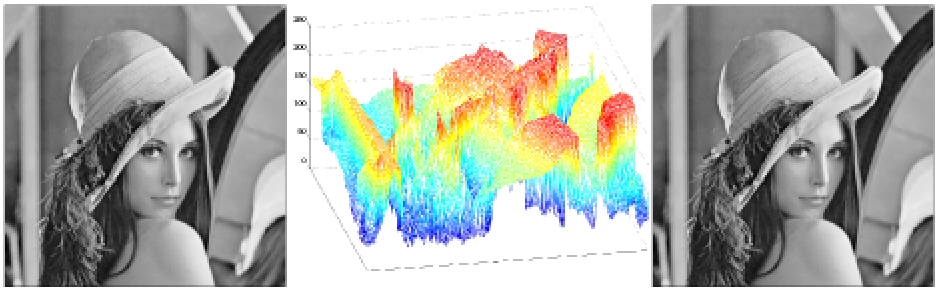
Subfigures

The Subsequent Figure Does Not Have a Name

Lena

Lena as a signal

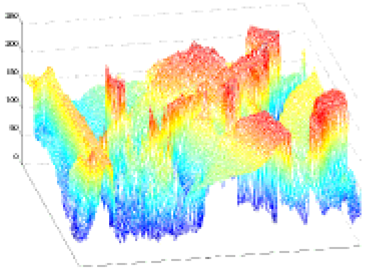
Lena again



On the left and right is the classic "Lena" image, which is used ubiquitously as a test image. It contains straight and curved lines, complicated texture, and a face. In the middle is a perspective display of the Lena image as a signal: a function of two spatial variables. The colors merely help show what signal values are about the same size. In this image, signal values range between 0 and 255; why is that?

This example is the same as the last, except that `orient='vertical'`. So, the subfigures should be on top of each other.

Lena



Lena as a signal

Tables

Here's is a very simple example of a table, similiar to ones seen previously in this module.

	Year		
Item	**1998**	**1999**	**2000**
Income	12,293	12,398	14,298
Expenditure	293	398	298
Surplus	12000	12000	14000

Budget - Comparison by yearThis caption put in to test styling.

Quotes and Foreign Terms

Two new elements added in CNXML 0.5 are **quote** and **foreign**. A **quote** of type "block" might look like this: "I define the **real** as that which holds its characters on such a tenure that it makes not the slightest difference what any man or men may have **thought** them to be, or ever will have **thought** them to be, here using thought to include, imagining, opining, and willing (as long as forcible **means** are not used); but the real thing's characters will remain absolutely untouched. --C.S. Peirce, *Collected Papers* 6.495." Anyone involved in interface design and usability should consider Einstein's axiom (shown here in a **quote** of type 'inline'): "Everything should be made as simple as possible, but not one bit simpler".

In order to appear urbane and literate, we should liberally sprinkle our discourse with foreign terms and phrases, such as Fahrvergnügen and sang-froid.

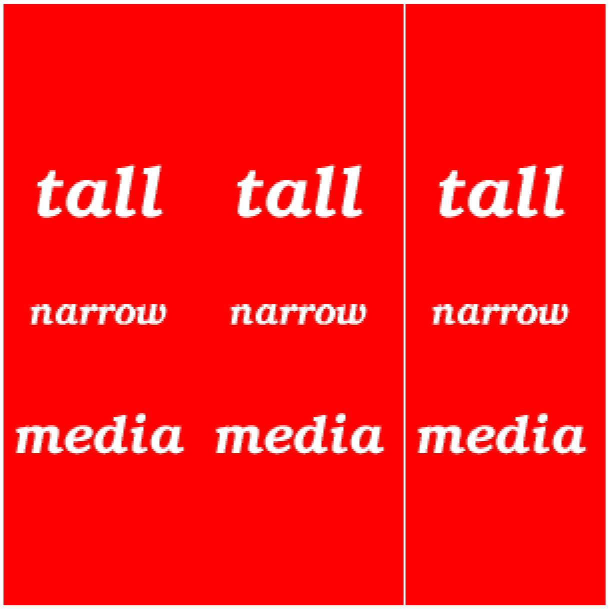
Note: It is possible to include a block quote inside a note, with **foreign** elements interspersed:

> "Freundschaftsbezeigungen" seems to be "Friendship demonstrations," which is only a foolish and clumsy way of saying "demonstrations of friendship." "Unabhängigkeitserklärungen" seems to be "Independencedeclarations," which is no improvement upon "Declarations of Independence," so far as I can see. "Generalstaatsverordnetenversammlungen" seems to be "General-statesrepresentativesmeetings," as nearly as I can get at it -- a mere rhythmical, gushy euphuism for "meetings of the legislature," I judge. --Mark Twain, *"The Awful German Language"*

Conclusion

For more information on all of the CNXML used in this module, please refer to the [CNXML Spec](#).

72 dpi	150 dpi	300 dpi



An example of how media that appear to be the same size on screen can be a different size in print due to resolution.

Glossary

glossary

A listing of terms together with their meanings.

[Misses](#)

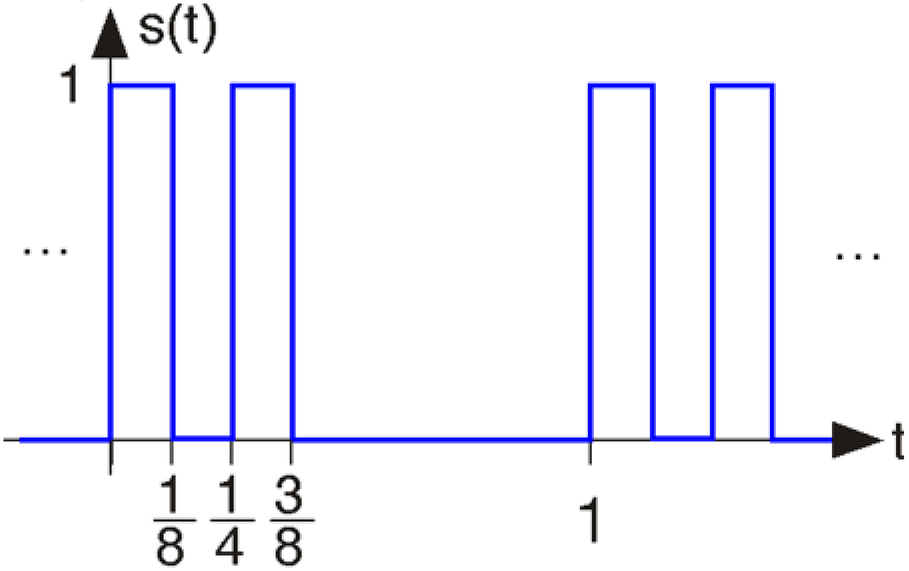
Example:
This very glossary.

Frequency Domain Problems (ticket #6364)
Problems dealing with Fourier Series.

Simple Fourier Series

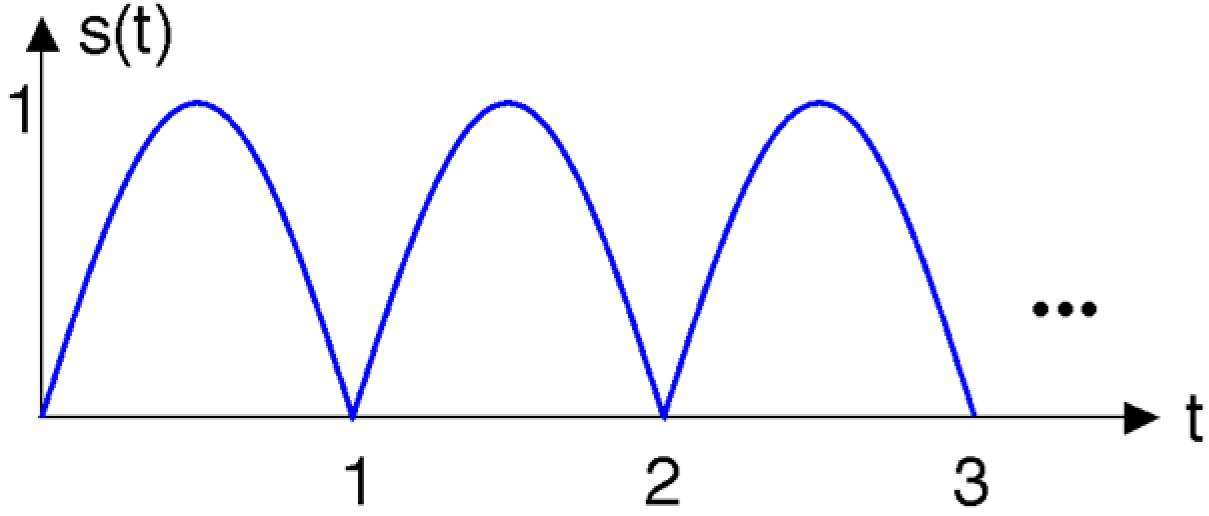
Find the complex Fourier series representations of the following signals without explicitly calculating Fourier integrals. What is the signal's period in each case?

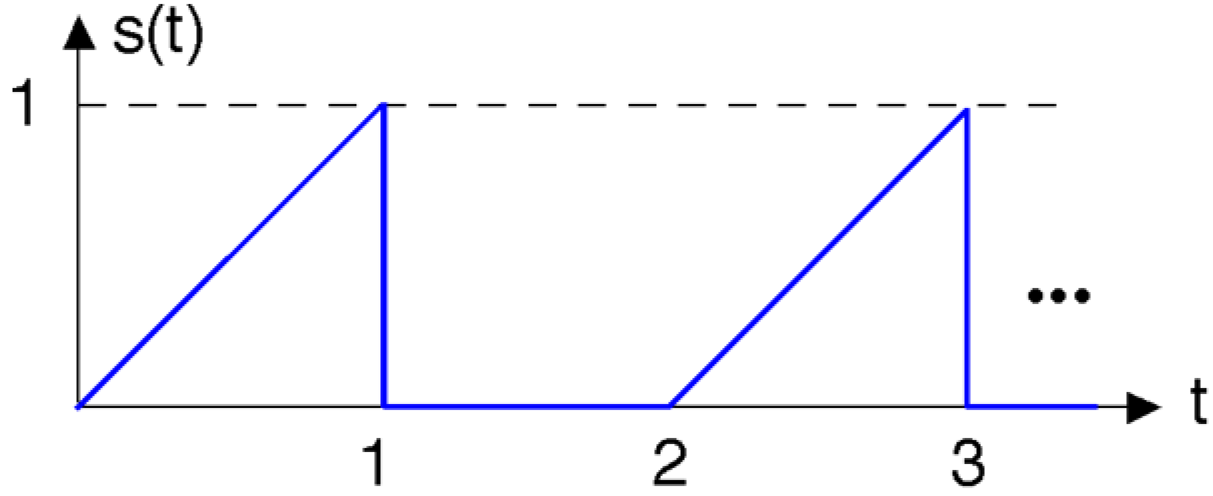
1. $s(t) = \sin(t)$
2. $s(t) = \sin^2(t)$
3. $s(t) = \cos(t) + 2\cos(2t)$
4. $s(t) = \cos(2t)\cos(t)$
5. $s(t) = \cos\left(10\pi t + \frac{\pi}{6}\right)(1 + \cos(2\pi t))$
6. $s(t)$ given by the depicted waveform.

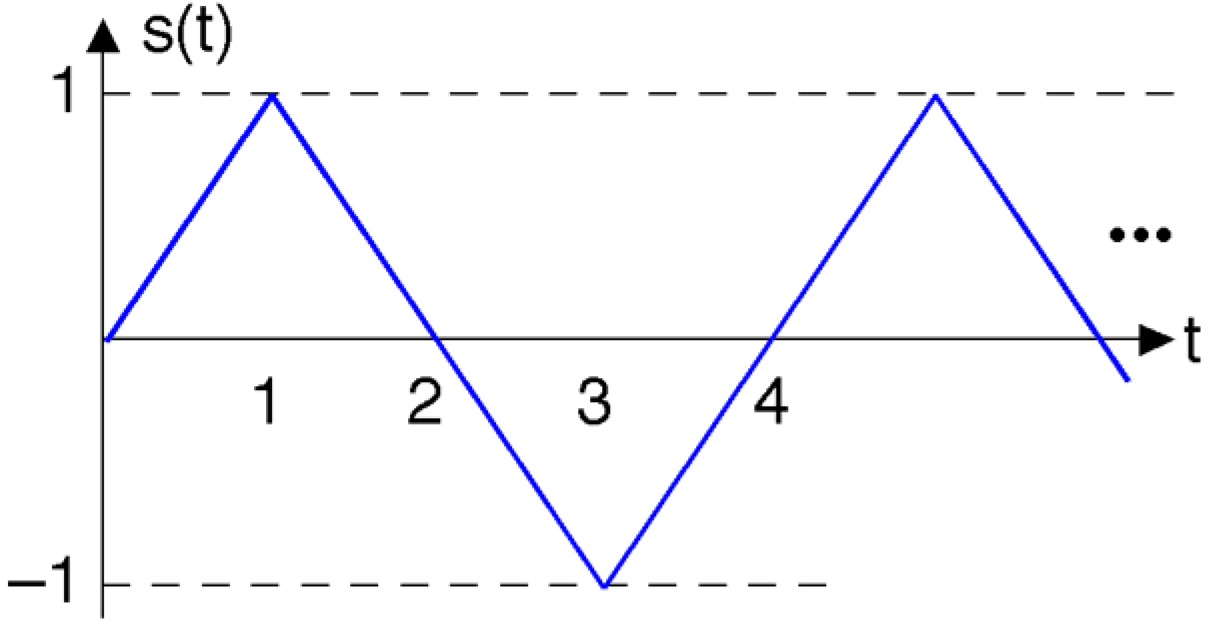


Fourier Series

Find the Fourier series representation for the following periodic signals. For the third signal, find the complex Fourier series for the triangle wave **without** performing the usual Fourier integrals. Hint: How is this signal related to one for which you already have the series?

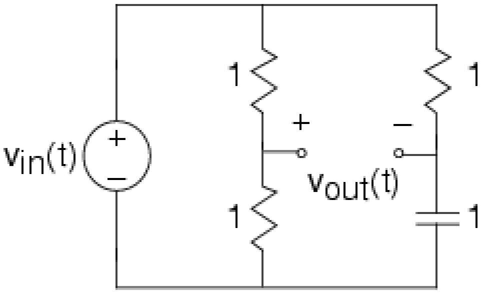

s(t)
1
…
t
1
2
3


s(t)
1
…
t
1
2
3


s(t)
1
…
t
1
2
3
4
–1

Phase Distortion

We can learn about phase distortion by returning to circuits and investigate the following circuit.



1. Find this filter's transfer function.
2. Find the magnitude and phase of this transfer function. How would you characterize this circuit?
3. Let $v_{\text{in}}(t)$ be a square-wave of period T. What is the Fourier series for the output voltage?
4. Use Matlab to find the output's waveform for the cases $T = 0.01$ and $T = 2$. What value of T delineates the two kinds of results you found? The software in `fourier2.m` might be useful.
5. Instead of the depicted circuit, the square wave is passed through a system that delays its input, which applies a linear phase shift to the signal's spectrum. Let the delay τ be $\frac{T}{4}$. Use the transfer function of a delay to compute using Matlab the Fourier series of the output. Show that the square wave is indeed delayed.

Approximating Periodic Signals

Often, we want to approximate a reference signal by a somewhat simpler signal. To assess the quality of an approximation, the most frequently used error measure is the mean-squared error. For a periodic signal $s(t)$,

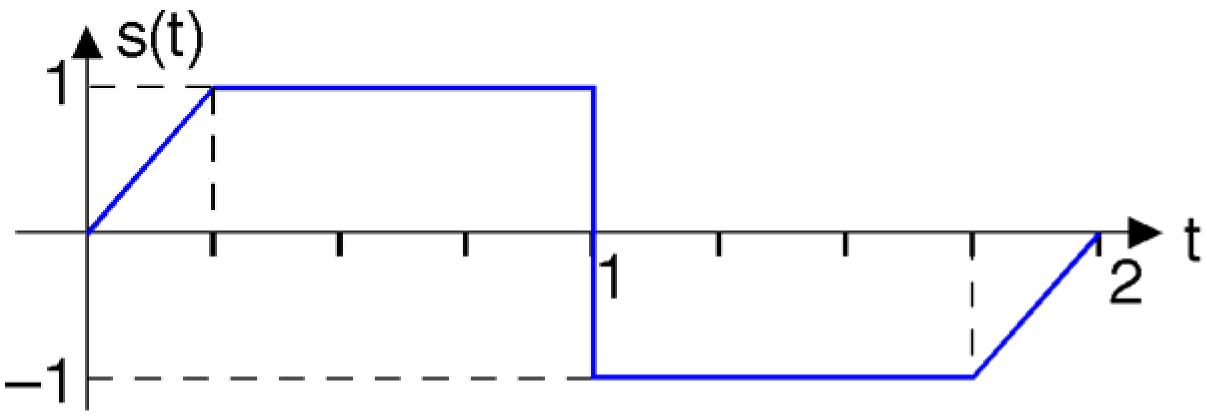
$$\varepsilon^2 = \frac{1}{T}\int_0^T \left(s(t) - \tilde{s}(t)\right)^2 \,\mathrm{d}\, t$$

where $s(t)$ is the reference signal and $\tilde{s}(t)$ its approximation. One convenient way of finding approximations for periodic signals is to truncate their Fourier series.

$$\tilde{s}(t) = \sum_{k=-K}^{K} c_k e^{i\frac{2\pi k}{T}t}$$

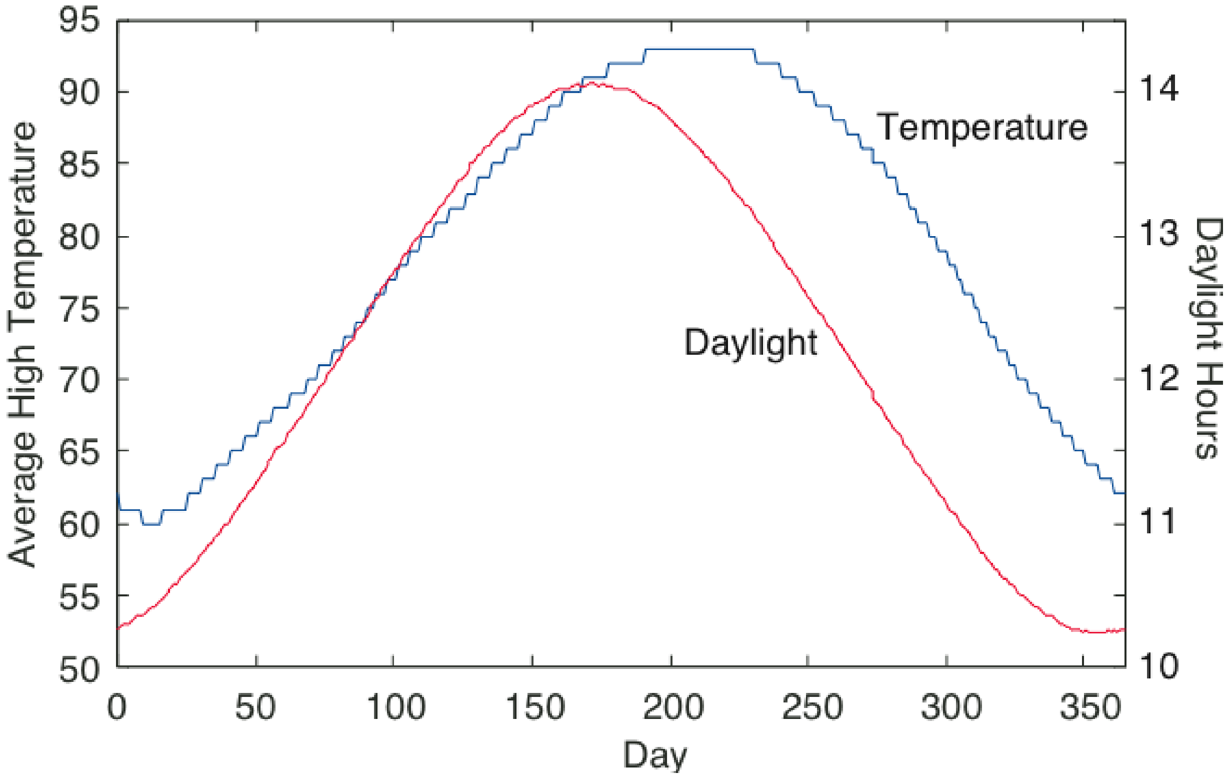
The point of this problem is to analyze whether this approach is the best (i.e., always minimizes the mean-squared error).

1. Find a frequency-domain expression for the approximation error when we use the truncated Fourier series as the approximation.
2. Instead of truncating the series, let's generalize the nature of the approximation to including any set of $2K+1$ terms: We'll always include the c_0 and the negative indexed term corresponding to c_k. What selection of terms minimizes the mean-squared error? Find an expression for the mean-squared error resulting from your choice.
3. Find the Fourier series for the depicted signal. Use Matlab to find the truncated approximation and best approximation involving two terms. Plot the mean-squared error as a function of K for both approximations.



Long, Hot Days

The daily temperature is a consequence of several effects, one of them being the sun's heating. If this were the dominant effect, then daily temperatures would be proportional to the number of daylight hours. The plot shows that the average daily high temperature does **not** behave that way.



In this problem, we want to understand the temperature component of our environment using Fourier series and linear system theory. The file `temperature.mat` contains these data (daylight hours in the first row, corresponding average daily highs in the second) for Houston, Texas.

1. Let the length of day serve as the sole input to a system having an output equal to the average daily temperature. Examining the plots of input and output, would you say that the system is linear or not? How did you reach you conclusion?
2. Find the first five terms (c_0, ... , c_4) of the complex Fourier series for each signal. Use the following formula that approximates the integral required to find the Fourier coefficients.

$$c_k = \frac{1}{366} \sum_{n=0}^{366} s(n) e^{-\left(i \frac{2\pi nk}{366}\right)}$$

3. What is the harmonic distortion in the two signals? Exclude c_0 from this calculation.
4. Because the harmonic distortion is small, let's concentrate only on the first harmonic. What is the phase shift between input and output signals?
5. Find the transfer function of the simplest possible linear model that would describe the data. Characterize and interpret the structure of this model. In particular, give a physical explanation for the phase shift.
6. Predict what the output would be if the model had no phase shift. Would days be hotter? If so, by how much?

Fourier Transform Pairs

Find the Fourier or inverse Fourier transform of the following.

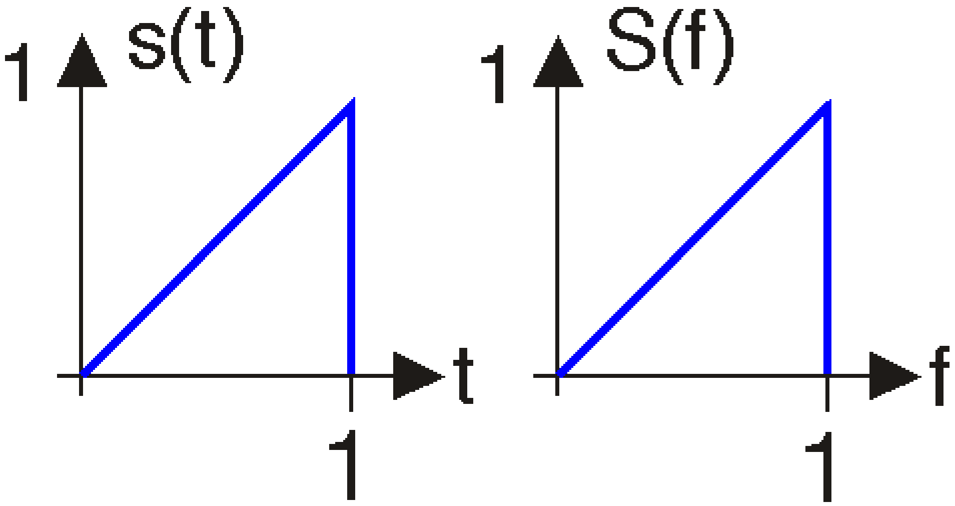
1. $\forall t : \left(x(t) = e^{-(a|t|)}\right)$
2. $x(t) = te^{-(at)}u(t)$
3. $X(f) = \begin{cases} 1 \text{ if } & |f| < W \\ 0 \text{ if } & |f| > W \end{cases}$
4. $x(t) = e^{-(at)} \cos(2\pi f_0 t) u(t)$

Duality in Fourier Transforms

"Duality" means that the Fourier transform and the inverse Fourier transform are very similar. Consequently, the waveform $s(t)$ in the time domain and the spectrum $s(f)$ have a Fourier transform and an inverse Fourier transform, respectively, that are very similar.

1. Calculate the Fourier transform of the signal shown below.
2. Calculate the inverse Fourier transform of the spectrum shown below.
3. How are these answers related? What is the general relationship between the Fourier transform of $s(t)$ and the inverse transform of

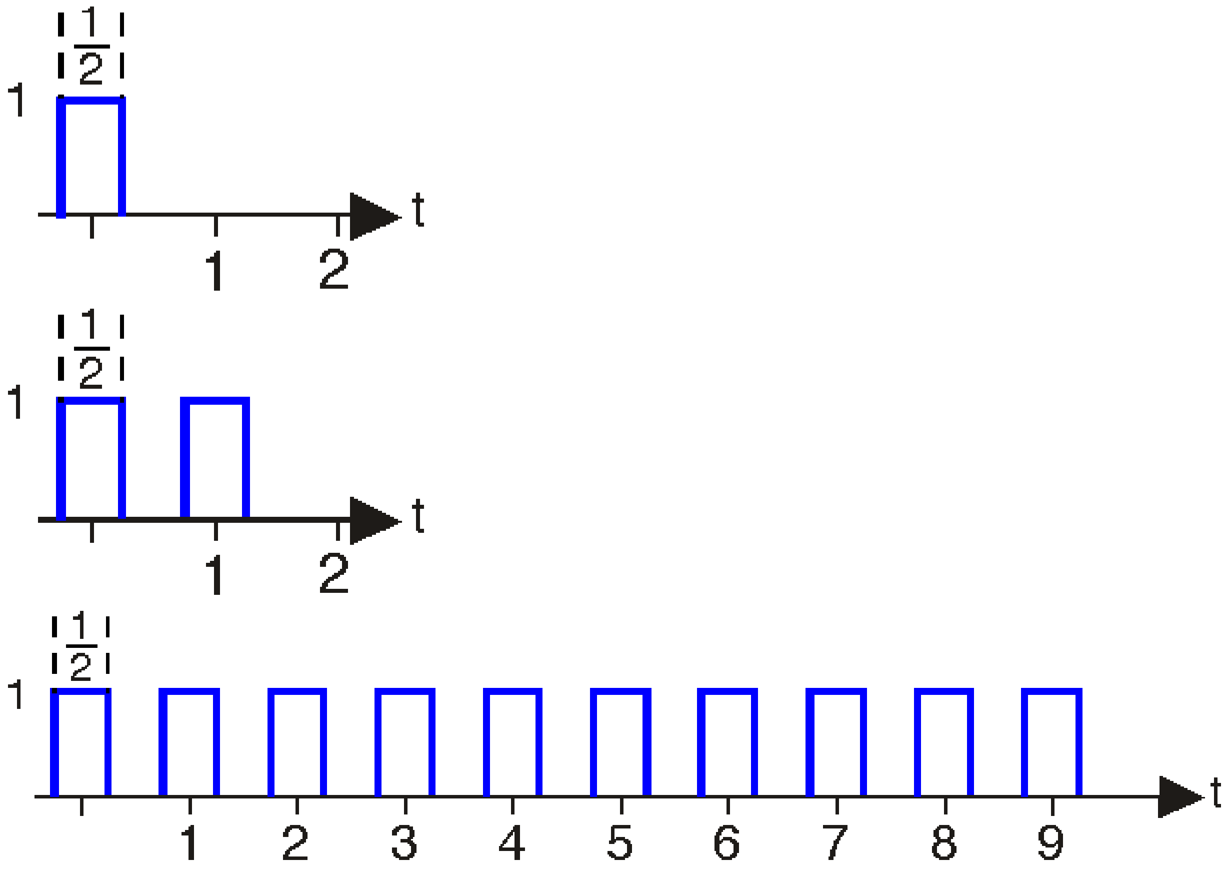
$s(f)$?



Spectra of Pulse Sequences

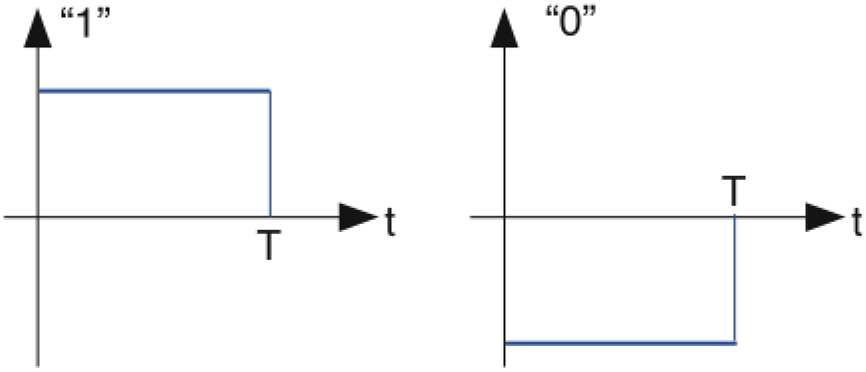
Pulse sequences occur often in digital communication and in other fields as well. What are their spectral properties?

1. Calculate the Fourier transform of the single pulse shown below.
2. Calculate the Fourier transform of the two-pulse sequence shown below.
3. Calculate the Fourier transform for the **ten**-pulse sequence shown in below. You should look for a general expression that holds for sequences of any length.
4. Using Matlab, plot the magnitudes of the three spectra. Describe how the spectra change as the number of repeated pulses increases.



Spectra of Digital Communication Signals

One way to represent bits with signals is shown in [link]. If the value of a bit is a “1”, it is represented by a positive pulse of duration T. If it is a “0”, it is represented by a negative pulse of the same duration. To represent a sequence of bits, the appropriately chosen pulses are placed one after the other.



1. What is the spectrum of the waveform that represents the alternating bit sequence "...01010101..."?
2. This signal's bandwidth is defined to be the frequency range over which 90% of the power is contained. What is this signal's bandwidth?
3. Suppose the bit sequence becomes "...00110011...". Now what is the bandwidth?

Lowpass Filtering a Square Wave

Let a square wave (period T) serve as the input to a first-order lowpass system constructed as a RC filter. We want to derive an expression for the time-domain response of the filter to this input.

1. First, consider the response of the filter to a simple pulse, having unit amplitude and width $\frac{T}{2}$. Derive an expression for the filter's output to this pulse.
2. Noting that the square wave is a superposition of a sequence of these pulses, what is the filter's response to the square wave?
3. The nature of this response should change as the relation between the square wave's period and the filter's cutoff frequency change. How long must the period be so that the response does **not** achieve a relatively constant value between transitions in the square wave? What is the relation of the filter's cutoff frequency to the square wave's spectrum in this case?

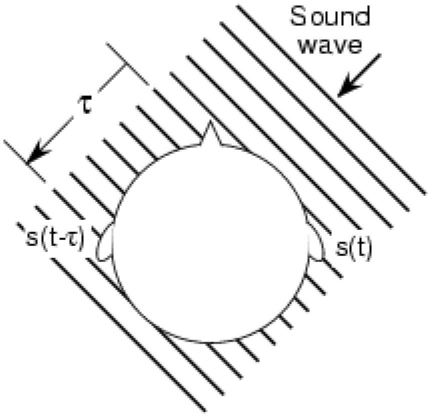
Mathematics with Circuits

Simple circuits can implement simple mathematical operations, such as integration and differentiation. We want to develop an active circuit (it contains an op-amp) having an output that is proportional to the integral of its input. For example, you could use an integrator in a car to determine distance traveled from the speedometer.

1. What is the transfer function of an integrator?
2. Find an op-amp circuit so that its voltage output is proportional to the integral of its input for all signals.

Where is that sound coming from?

We determine where sound is coming from because we have two ears and a brain. Sound travels at a relatively slow speed and our brain uses the fact that sound will arrive at one ear before the other. As shown here, a sound coming from the right arrives at the left ear τ seconds after it arrives at the right ear.



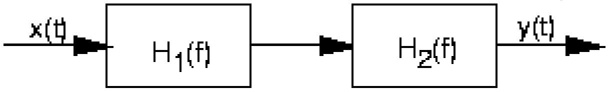
Once the brain finds this propagation delay, it can determine the sound direction. In an attempt to model what the brain might do, RU signal processors want to design an **optimal** system that delays each ear's signal by some amount then adds them together. Δ_l and Δ_r are the delays applied to the left and right signals respectively. The idea is to determine the delay values according to some criterion that is based on what is measured by the two ears.

1. What is the transfer function between the sound signal $s(t)$ and the processor output $y(t)$?
2. One way of determining the delay τ is to choose Δ_l and Δ_r to maximize the power in $y(t)$. How are these maximum-power processing delays related to τ?

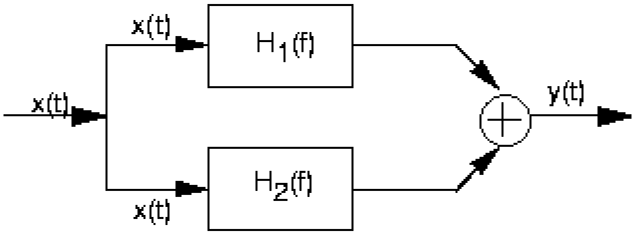
Arrangements of Systems

Architecting a system of modular components means arranging them in various configurations to achieve some overall input-output relation. For each of the following, determine the overall transfer function between $x(t)$ and $y(t)$.

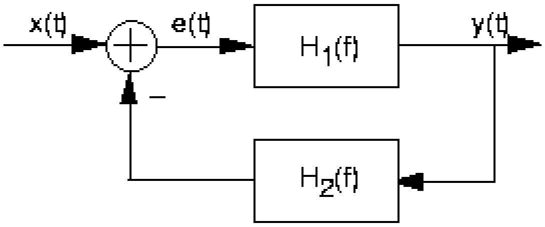
system a



system b



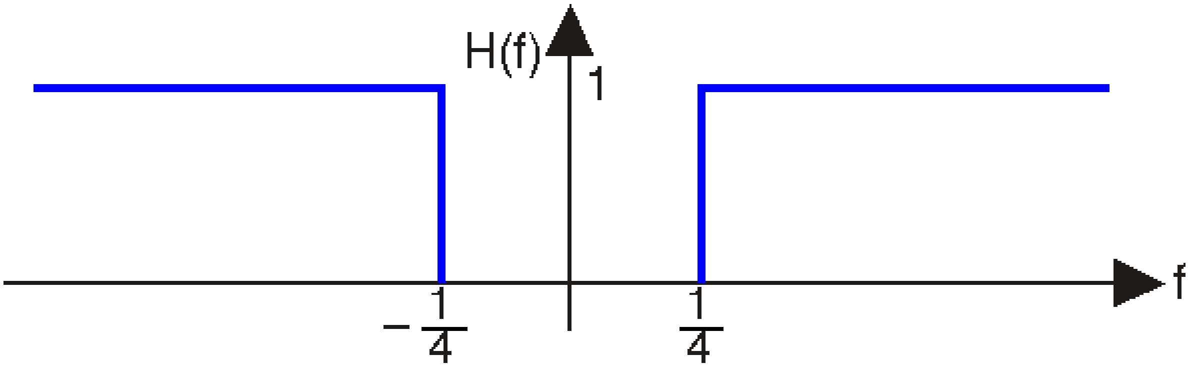
system c



The overall transfer function for the cascade (first depicted system) is particularly interesting. What does it say about the effect of the ordering of linear, time-invariant systems in a cascade?

Filtering

Let the signal $s(t) = \frac{\sin(\pi t)}{\pi t}$ be the input to a linear, time-invariant filter having the transfer function shown below. Find the expression for $y(t)$, the filter's output.



Circuits Filter!

A unit-amplitude pulse with duration of one second serves as the input to an RC-circuit having transfer function

$$H(f) = \frac{i2\pi f}{4 + i2\pi f}$$

1. How would you categorize this transfer function: lowpass, highpass, bandpass, other?
2. Find a circuit that corresponds to this transfer function.
3. Find an expression for the filter's output.

Reverberation

Reverberation corresponds to adding to a signal its delayed version.

1. Assuming τ represents the delay, what is the input-output relation for a reverberation system? Is the system linear and time-invariant? If so, find the transfer function; if not, what linearity or time-invariance criterion does reverberation violate.

2. A music group known as the ROwls is having trouble selling its recordings. The record company's engineer gets the idea of applying different delay to the low and high frequencies and adding the result to create a new musical effect. Thus, the ROwls' audio would be separated into two parts (one less than the frequency f_0, the other greater than f_0), these would be delayed by τ_l and τ_h respectively, and the resulting signals added. Draw a block diagram for this new audio processing system, showing its various components.
3. How does the magnitude of the system's transfer function depend on the two delays?

Echoes in Telephone Systems

A frequently encountered problem in telephones is echo. Here, because of acoustic coupling between the ear piece and microphone in the handset, what you hear is also sent to the person talking. That person thus not only hears you, but also hears her own speech delayed (because of propagation delay over the telephone network) and attenuated (the acoustic coupling gain is less than one). Furthermore, the same problem applies to you as well: The acoustic coupling occurs in her handset as well as yours.

1. Develop a block diagram that describes this situation.
2. Find the transfer function between your voice and what the listener hears.
3. Each telephone contains a system for reducing echoes using electrical means. What simple system could null the echoes?

Effective Drug Delivery

In most patients, it takes time for the concentration of an administered drug to achieve a constant level in the blood stream. Typically, if the drug concentration in the patient's intravenous line is $C_d u(t)$, the concentration in the patient's blood stream is $C_p \left(1 - e^{-(at)}\right) u(t)$.

1. Assuming the relationship between drug concentration in the patient's drug and the delivered concentration can be described as a linear, time-

invariant system, what is the transfer function?

2. Sometimes, the drug delivery system goes awry and delivers drugs with little control. What would the patient's drug concentration be if the delivered concentration were a ramp? More precisely, if it were $C_d tu(t)$?
3. A clever doctor wants to have the flexibility to slow down or speed up the patient's drug concentration. In other words, the concentration is to be $C_p\left(1-e^{-(bt)}\right)u(t)$, with b bigger or smaller than a. How should the delivered drug concentration signal be changed to achieve this concentration profile?

Catching Speeders with Radar

RU Electronics has been contracted to design a Doppler radar system. Radar transmitters emit a signal that bounces off any conducting object. Signal differences between what is sent and the radar return is processed and features of interest extracted. In **Doppler** systems, the object's speed along the direction of the radar beam is the feature the design must extract. The transmitted signal is a sinsusoid: $x(t)=A\cos(2\pi f_c t)$. The measured return signal equals $B\cos(2\pi\left((f_c+\Delta \mathrm{f})t+\phi\right))$, where the Doppler offset frequency $\Delta \mathrm{f}$ equals $10v$, where v is the car's velocity coming toward the transmitter.

1. Design a system that uses the transmitted and return signals as inputs and produces $\Delta \mathrm{f}$.
2. One problem with designs based on overly simplistic design goals is that they are sensitive to unmodeled assumptions. How would you change your design, if at all, so that whether the car is going away or toward the transmitter could be determined?
3. Suppose two objects traveling different speeds provide returns. How would you change your design, if at all, to accomodate multiple returns?

Demodulating an AM Signal

Let $m(t)$ denote the signal that has been amplitude modulated.

$$x(t) = A\,(1 + m(t))\sin(2\pi f_c t)$$

Radio stations try to restrict the amplitude of the signal $m(t)$ so that it is less than one in magnitude. The frequency f_c is very large compared to the frequency content of the signal. What we are concerned about here is not transmission, but reception.

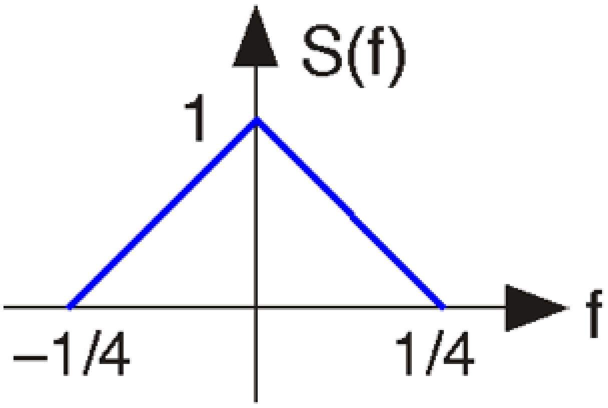
1. The so-called coherent demodulator simply multiplies the signal $x(t)$ by a sinusoid having the same frequency as the carrier and lowpass filters the result. Analyze this receiver and show that it works. Assume the lowpass filter is ideal.
2. One issue in coherent reception is the phase of the sinusoid used by the receiver relative to that used by the transmitter. Assuming that the sinusoid of the receiver has a phase φ, how does the output depend on φ? What is the worst possible value for this phase?
3. The incoherent receiver is more commonly used because of the phase sensitivity problem inherent in coherent reception. Here, the receiver full-wave rectifies the received signal and lowpass filters the result (again ideally). Analyze this receiver. Does its output differ from that of the coherent receiver in a significant way?

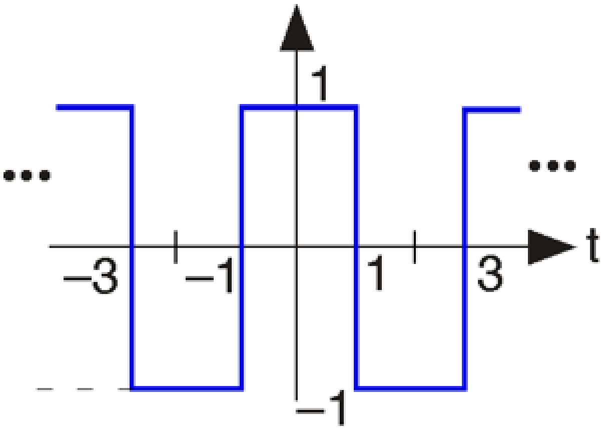
Unusual Amplitude Modulation

We want to send a band-limited signal having the depicted spectrum with amplitude modulation in the usual way. I.B. Different suggests using the square-wave carrier shown below. Well, it is different, but his friends wonder if any technique can demodulate it.

1. Find an expression for $X(f)$, the Fourier transform of the modulated signal.
2. Sketch the magnitude of $X(f)$, being careful to label important magnitudes and frequencies.
3. What demodulation technique obviously works?
4. I.B. challenges three of his friends to demodulate $x(t)$ some other way. One friend suggests modulating $x(t)$ with $\cos\left(\frac{\pi t}{2}\right)$, another wants to try modulating with $\cos(\pi t)$ and the third thinks $\cos\left(\frac{3\pi t}{2}\right)$ will work.

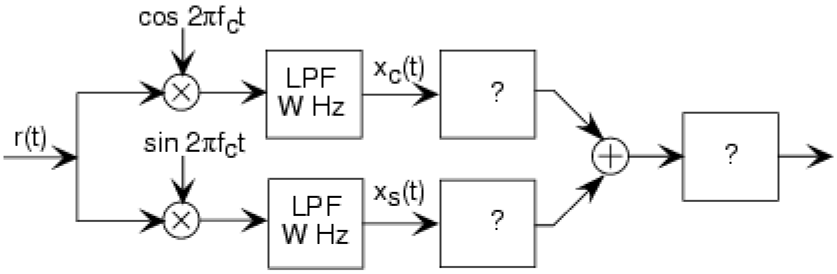
Sketch the magnitude of the Fourier transform of the signal each student's approach produces. Which student comes closest to recovering the original signal? Why?





Sammy Falls Asleep...

While sitting in ELEC 241 class, he falls asleep during a critical time when an AM receiver is being described. The received signal has the form $r(t) = A\,(1 + m(t))\cos(2\pi f_c t + \varphi)$ where the phase φ is unknown. The message signal is $m(t)$; it has a bandwidth of W Hz and a magnitude less than 1 ($|m(t)| < 1$). The phase φ is unknown. The instructor drew a diagram for a receiver on the board; Sammy slept through the description of what the unknown systems where.



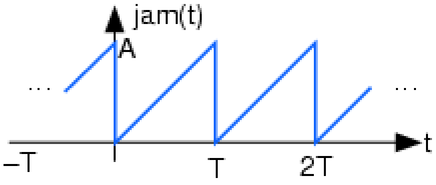
1. What are the signals $x_c(t)$ and $x_s(t)$?
2. What would you put in for the unknown systems that would guarantee that the final output contained the message regardless of the phase?

Note: Think of a trigonometric identity that would prove useful.

3. Sammy may have been asleep, but he can think of a far simpler receiver. What is it?

Jamming

Sid Richardson college decides to set up its own AM radio station KSRR. The resident electrical engineer decides that she can choose **any** carrier frequency and message bandwidth for the station. A rival college decides to **jam** its transmissions by transmitting a high-power signal that interferes with radios that try to receive KSRR. The jamming signal jam (t) is what is known as a **sawtooth** wave (depicted in [link]) having a period known to KSRR's engineer.

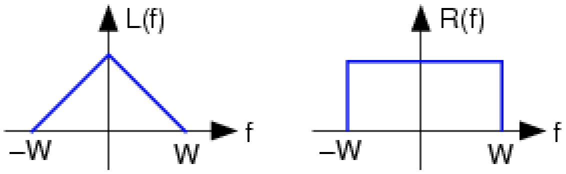


1. Find the spectrum of the jamming signal.

2. Can KSRR entirely circumvent the attempt to jam it by carefully choosing its carrier frequency and transmission bandwidth? If so, find the station's carrier frequency and transmission bandwidth in terms of T, the period of the jamming signal; if not, show why not.

AM Stereo

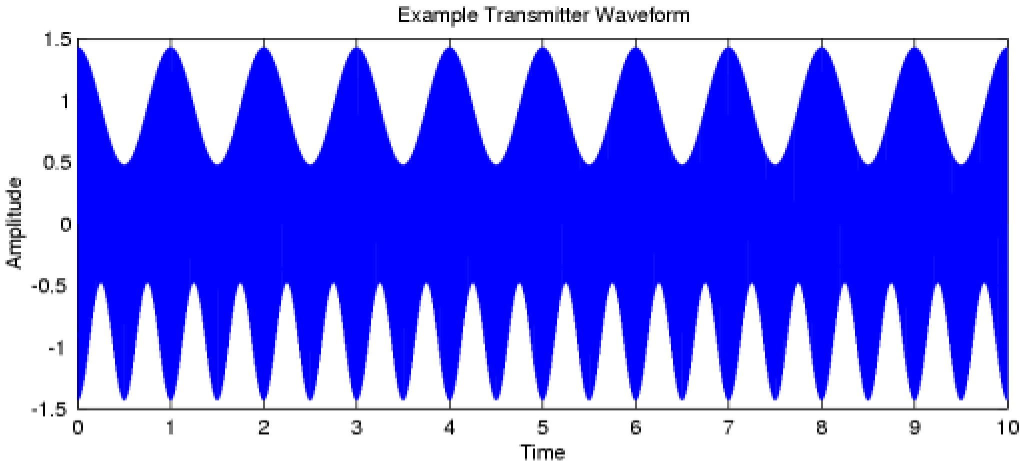
A stereophonic signal consists of a "left" signal $l(t)$ and a "right" signal $r(t)$ that conveys sounds coming from an orchestra's left and right sides, respectively. To transmit these two signals simultaneously, the transmitter first forms the sum signal $s_{+}(t) = l(t) + r(t)$ and the difference signal $s_{-}(t) = l(t) - r(t)$. Then, the transmitter amplitude-modulates the difference signal with a sinusoid having frequency $2W$, where W is the bandwidth of the left and right signals. The sum signal and the modulated difference signal are added, the sum amplitude-modulated to the radio station's carrier frequency f_c, and transmitted. Assume the spectra of the left and right signals are as shown.



1. What is the expression for the transmitted signal? Sketch its spectrum.
2. Show the block diagram of a stereo AM receiver that can yield the left and right signals as separate outputs.
3. What signal would be produced by a conventional coherent AM receiver that expects to receive a standard AM signal conveying a message signal having bandwidth W?

Novel AM Stereo Method

A clever engineer has submitted a patent for a new method for transmitting two signals **simultaneously** in the **same** transmission bandwidth as commercial AM radio. As shown, her approach is to modulate the positive portion of the carrier with one signal and the negative portion with a second.



In detail the two message signals $m_1(t)$ and $m_2(t)$ are bandlimited to W Hz and have maximal amplitudes equal to 1. The carrier has a frequency f_c much greater than W. The transmitted signal $x(t)$ is given by

$$x(t) = \begin{cases} A\left(1 + a m_1(t)\right)\sin(2\pi f_c t) & \text{if} \quad \sin(2\pi f_c t) \geq 0 \\ A\left(1 + a m_2(t)\right)\sin(2\pi f_c t) & \text{if} \quad \sin(2\pi f_c t) < 0 \end{cases}$$

In all cases, $0 < a < 1$. The plot shows the transmitted signal when the messages are sinusoids: $m_1(t) = \sin(2\pi f_m t)$ and $m_2(t) = \sin(2\pi 2 f_m t)$ where $2f_m < W$. You, as the patent examiner, must determine whether the scheme meets its claims and is useful.

1. Provide a more concise expression for the transmitted signal $x(t)$ than given above.
2. What is the receiver for this scheme? It would yield both $m_1(t)$ and $m_2(t)$ from $x(t)$.
3. Find the spectrum of the positive portion of the transmitted signal.
4. Determine whether this scheme satisfies the design criteria, allowing you to grant the patent. Explain your reasoning.

A Radical Radio Idea

An ELEC 241 student has the bright idea of using a square wave instead of a sinusoid as an AM carrier. The transmitted signal would have the form

$$x(t) = A\left(1 + m(t)\right)\mathrm{sq}_{T}\left(t\right)$$

where the message signal $m(t)$ would be amplitude-limited: $|m(t)| < 1$

1. Assuming the message signal is lowpass and has a bandwidth of W Hz, what values for the square wave's period T are feasible. In other words, do some combinations of W and T prevent reception?
2. Assuming reception is possible, can **standard** radios receive this innovative AM transmission? If so, show how a coherent receiver could demodulate it; if not, show how the coherent receiver's output would be corrupted. Assume that the message bandwidth $W = 5$ kHz.

Secret Communication

An amplitude-modulated secret message $m(t)$ has the following form.

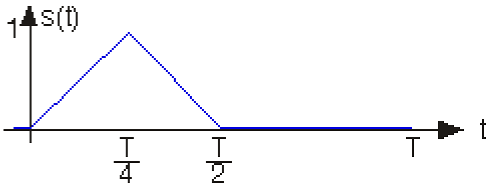
$$r(t) = A\left(1 + m(t)\right)\cos(2\pi\left(f_c + f_0\right)t)$$

The message signal has a bandwidth of W Hz and a magnitude less than 1 ($|m(t)| < 1$). The idea is to offset the carrier frequency by f_0 Hz from standard radio carrier frequencies. Thus, "off-the-shelf" coherent demodulators would assume the carrier frequency has f_c Hz. Here, $f_0 < W$.

1. Sketch the spectrum of the demodulated signal produced by a coherent demodulator tuned to f_c Hz.
2. Will this demodulated signal be a “scrambled” version of the original? If so, how so; if not, why not?
3. Can you develop a receiver that can demodulate the message without knowing the offset frequency f_c?

Signal Scrambling

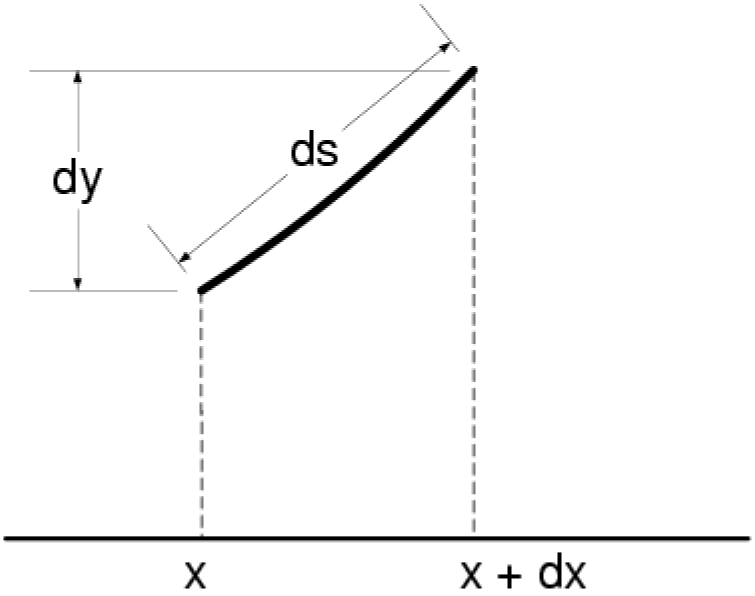
An excited inventor announces the discovery of a way of using analog technology to render music unlistenable without knowing the secret recovery method. The idea is to modulate the bandlimited message $m(t)$ by a special periodic signal $s(t)$ that is zero during half of its period, which renders the message unlistenable and superficially, at least, unrecoverable ([link]).



1. What is the Fourier series for the periodic signal?
2. What are the restrictions on the period T so that the message signal can be recovered from $m(t)s(t)$?
3. ELEC 241 students think they have "broken" the inventor's scheme and are going to announce it to the world. How would they recover the original message **without** having detailed knowledge of the modulating signal?

Energy in a mechanical wave

Energy Transport



Lets calculate the energy in a wave of a string:

Consider a fragment of string so small it can be considered straight, as is shown in the figure

The the kinetic energy is $K = \frac{1}{2}m\mathrm{v}^2$ for the string fragment $m = \mu dx$

Note: Why is this dx and not ds? Lets consider that the string is not perturbed, then it is horizontal and has mass as given. When the string is perturbed it stretches a little bit - but the mass does not increase.

So we have

$$K = \frac{1}{2}\mu \mathbb{d}x \left(\frac{\partial y}{\partial t}\right)^2$$

and using this we can define the energy per unit length, ie. the kinetic energy density:

$$\frac{\mathbb{d}K}{\mathbb{d}x} = \frac{1}{2}\mu \left(\frac{\partial y}{\partial t}\right)^2$$

When the string segment is stretched from the length $\mathbb{d}x$ to the length $\mathbb{d}s$ an amount of work $= T(\mathbb{d}s - \mathbb{d}x)$ is done. This is equal to the potential energy stored in the stretched string segment. So the potential energy in this case is:

$$U = T(\mathbb{d}s - \mathbb{d}x)$$

Now

$$\begin{aligned} \mathbb{d}s &= (\mathbb{d}x^2 + \mathbb{d}y^2)^{1/2} \\ &= \mathbb{d}x \left[1 + \left(\tfrac{\partial y}{\partial x}\right)^2\right]^{1/2} \end{aligned}$$

Recall the binomial expansion

$$(1 + A)^n = 1 + nA + \frac{n(n-1)A^2}{2!} + \frac{n(n-1)(n-2)A^3}{3!} + \ldots$$

so

$$\mathbb{d}s \approx \mathbb{d}x + \frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2 \mathbb{d}x$$

$$U = T(\mathbb{d}s - \mathbb{d}x) \approx \frac{1}{2}T\left(\frac{\partial y}{\partial x}\right)^2 \mathbb{d}x$$

or the potential energy density

$$\frac{\mathrm{d}U}{\mathrm{d}x} = \frac{1}{2}T\left(\frac{\partial y}{\partial x}\right)^2$$

To get the kinetic energy in a wavelength, lets start with

$$y = A\sin\left(\frac{2\pi x}{\lambda} - \omega t\right)$$

$$\frac{\partial y}{\partial t} = -\omega A\cos\left(\frac{2\pi x}{\lambda} - \omega t\right)$$

Lets evaluate it at time 0.

$$\left.\frac{\partial y}{\partial t}\right|_{t=0} = -\omega A\cos\left(\frac{2\pi x}{\lambda}\right)$$

so

$$\frac{\mathrm{d}K}{\mathrm{d}x} = \frac{1}{2}\mu\omega^2 A^2\cos^2\left(\frac{2\pi x}{\lambda}\right)$$

now integrate

$$\begin{gathered} K = \int_0^\lambda \tfrac{\mathrm{d}K}{\mathrm{d}x}\mathrm{d}x \\ = \tfrac{1}{2}\mu\omega^2 A^2\int_0^\lambda \cos^2\left(\tfrac{2\pi x}{\lambda}\right)\mathrm{d}x \end{gathered}$$

In order to do this integral we use the following trig identity:

$$\cos^2 A = \frac{\cos 2A + 1}{2}$$

so we get

$$\begin{gathered} K = \tfrac{1}{2}\mu\omega^2 A^2\left[\tfrac{x}{2} + \tfrac{\lambda}{8\pi}\sin\tfrac{4\pi x}{\lambda}\right]\Big|_{x=0}^{\lambda} \\ = \tfrac{1}{4}\mu\lambda\omega^2 A^2 \end{gathered}$$

In similar fashion the potential energy can be found to be

$$U = \frac{1}{4}\mu\lambda\omega^2 A^2.$$

Deriving this will be assigned as a homework problem

So

$$E = K + U = \frac{1}{2}\mu\lambda\omega^2 A^2$$

Power

$$\begin{aligned} P &= \frac{\Delta E}{\Delta t} = \frac{\frac{1}{2}\mu\lambda\omega^2 A^2}{\tau} \\ &= \frac{1}{2}\mu\omega^2 A^2 \mathbf{v} \end{aligned}$$

Where I have used $\tau = 1/\nu$ and $\lambda\nu =$v thus $\tau = \lambda/\mathbf{v}$

Árvores binárias de pesquisa

Árvores de pesquisa

Ah! este parágrafo é necessário para que a seção tenha algo a dizer. Vejam! e a fonte é diferente!

Suponha que você tenha um conjunto de elementos a armazenar no seu computador, de forma que possa realizar inúmeras buscas sobre esses elementos. Suponha agora que você utilize, como estrutura de dados para armazenar esses elementos, uma lista encadeada. Como vocês devem lembrar, a complexidade de busca em uma lista encadeada é (no pior caso) O(n), onde n é o número de elementos da lista. Isso ocorre porque precisamos percorrer a lista, no pior caso (caso onde o elemento a ser procurado se encontra no final da lista), do início (cabeça da lista) até o final. E se pudéssemos dar “saltos” nessa procura?

Claro que podemos dar “saltos” em uma procura. Alias, essa é a forma natural de otimizar processos. Para melhor visualizar esse processo, imagine uma busca de um número no intervalo de 0 a 100. Ao invés de iniciar sua busca no 0 e se estender até o 100, você pode testar o médio, 50, e verificar se o número a ser encontrado é maior ou menor. Se for menor, podemos então desprezar todos os números acima de 50 e recomeçar a busca no intervalo de 0 a 50. Esse procedimento deverá se repetir até que o número correspondente seja encontrado. Esse procedimento gerou implicitamente uma árvore de busca, conforme ilustrado na figura .

`para todo vértice além-mar faça com`

Responda a pergunta do exercício para melhor fixar seu conteúdo.

Glossary

Árvore binária estrita
    Árvore binária onde cada nó não-folha possui os dois filhos

Hypothesis Testing of Single Mean and Single Proportion: Examples (ticket #6053)

This module provides examples of Hypothesis Testing of a Single Mean and a Single Proportion as a part of the Collaborative Statistics collection (col10522) by Barbara Illowsky and Susan Dean.

Example:
Exercise:

Problem:

Jeffrey, as an eight-year old, **established a mean time of 16.43 seconds** for swimming the 25-yard freestyle, with a **standard deviation of 0.8 seconds**. His dad, Frank, thought that Jeffrey could swim the 25-yard freestyle faster by using goggles. Frank bought Jeffrey a new pair of expensive goggles and timed Jeffrey for **15 25-yard freestyle swims**. For the 15 swims, **Jeffrey's mean time was 16 seconds. Frank thought that the goggles helped Jeffrey to swim faster than the 16.43 seconds.** Conduct a hypothesis test using a preset $\alpha = 0.05$. Assume that the swim times for the 25-yard freestyle are normal.

Solution:

Set up the Hypothesis Test:

Since the problem is about a mean, this is a **test of a single population mean**.

H_{o}: $\mu = 16.43$ $\qquad$ H_a: $\mu < 16.43$

For Jeffrey to swim faster, his time will be less than 16.43 seconds. The "$<$" tells you this is left-tailed.

Determine the distribution needed:

Random variable: $\bar{X}$ = the mean time to swim the 25-yard freestyle.

Distribution for the test: X is normal (population **standard deviation** is known: $\sigma = 0.8$)

$X \sim N\left(\mu, \frac{\sigma_X}{\sqrt{n}}\right)$ Therefore, $X \sim N\left(16.43, \frac{0.8}{\sqrt{15}}\right)$

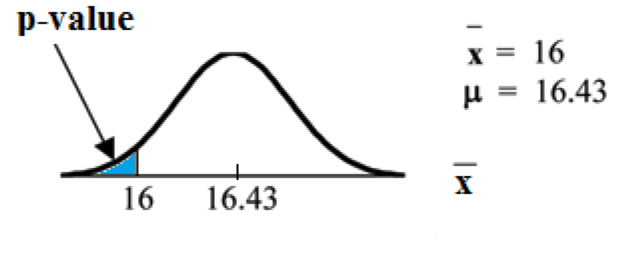
$\mu = 16.43$ comes from H_0 and not the data. $\sigma = 0.8$, and $n = 15$.

Calculate the p-value using the normal distribution for a mean:

p-value $= P\left(\bar{x} < 16\right) = 0.0187$ where the sample mean in the problem is given as 16.

p-value $= 0.0187$ (This is called the **actual level of significance**.) The p-value is the area to the left of the sample mean is given as 16.

Graph:



$\mu = 16.43$ comes from H_o. Our assumption is $\mu = 16.43$.

Interpretation of the p-value: If H_o is true, there is a 0.0187 probability (1.87%) that Jeffrey's mean time to swim the 25-yard freestyle is 16 seconds or less. Because a 1.87% chance is small, the mean time of 16 seconds or less is unlikely to have happened randomly. It is a rare event.

Compare α and the p-value:

$\alpha = 0.05$ p-value $= 0.0187$ $\alpha >$ p-value

Make a decision: Since $\alpha >$ p-value, reject H_o.

This means that you reject $\mu = 16.43$. In other words, you do not think Jeffrey swims the 25-yard freestyle in 16.43 seconds but faster with the new goggles.

Conclusion: At the 5% significance level, we conclude that Jeffrey swims faster using the new goggles. The sample data show there is sufficient evidence that Jeffrey's mean time to swim the 25-yard freestyle is less than 16.43 seconds.

The p-value can easily be calculated using the TI-83+ and the TI-84 calculators:

Press `STAT` and arrow over to `TESTS`. Press `1:Z-Test`. Arrow over to `Stats` and press `ENTER`. Arrow down and enter 16.43 for μ_0 (null hypothesis), .8 for σ, 16 for the sample mean, and 15 for n. Arrow down to μ: (alternate hypothesis) and arrow over to $<\mu_0$. Press `ENTER`. Arrow down to `Calculate` and press `ENTER`. The calculator not only calculates the p-value ($p = 0.0187$) but it also calculates the test statistic (z-score) for the sample mean. $\mu < 16.43$ is the alternate hypothesis. Do this set of instructions again except arrow to `Draw` (instead of `Calculate`). Press `ENTER`. A shaded graph appears with $z = -2.08$ (test statistic) and $p = 0.0187$ (p-value). Make sure when you use `Draw` that no other equations are highlighted in $Y =$ and the plots are turned off.

When the calculator does a Z-Test, the `Z-Test` function finds the p-value by doing a normal probability calculation using the **Central Limit Theorem**:

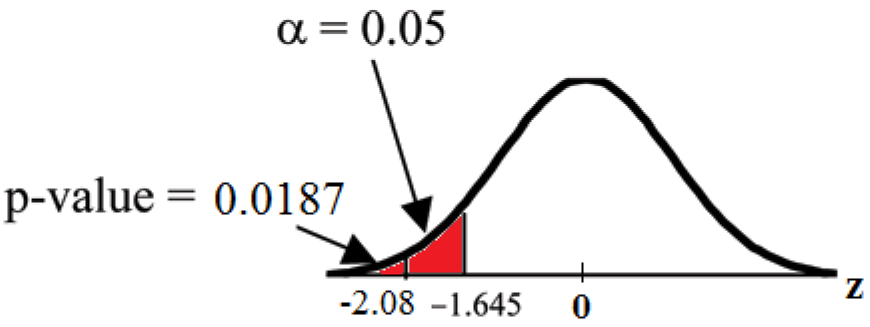
$P(x < 16\,) =$ `2nd DISTR normcdf` $\left(-10\ \hat{}\ 99, 16, 16.43, 0.8/\sqrt{15}\right)$.

The Type I and Type II errors for this problem are as follows:

The Type I error is to conclude that Jeffrey swims the 25-yard freestyle, on average, in less than 16.43 seconds when, in fact, he actually swims the 25-yard freestyle, on average, in 16.43 seconds. (Reject the null hypothesis when the null hypothesis is true.)

The Type II error is that there is not evidence to conclude that Jeffrey swims the 25-yard free-style, on average, in less than 16.43 seconds when, in fact, he actually does swim the 25-yard free-style, on average, in less than 16.43 seconds. (Do not reject the null hypothesis when the null hypothesis is false.)

Historical Note: The traditional way to compare the two probabilities, α and the p-value, is to compare the critical value (z-score from α) to the test statistic (z-score from data). The calculated test statistic for the p-value is -2.08. (From the Central Limit Theorem, the test statistic formula is $z = \frac{x-\mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)}$. For this problem, $x = 16$, $\mu_X = 16.43$ from the null hypothesis, $\sigma_X = 0.8$, and $n = 15$.) You can find the critical value for $\alpha = 0.05$ in the normal table (see **15.Tables** in the Table of Contents). The z-score for an area to the left equal to 0.05 is midway between -1.65 and -1.64 (0.05 is midway between 0.0505 and 0.0495). The z-score is -1.645. Since $-1.645 > -2.08$ (which demonstrates that $\alpha >$ p-value), reject H_o. Traditionally, the decision to reject or not reject was done in this way. Today, comparing the two probabilities α and the p-value is very common. For this problem, the p-value, 0.0187 is considerably smaller than α, 0.05. You can be confident about your decision to reject. The graph shows α, the p-value, and the test statistics and the critical value.



Example:
Exercise:

Problem:

A college football coach thought that his players could bench press a **mean weight of 275 pounds**. It is known that the **standard deviation is 55 pounds**. Three of his players thought that the mean weight was **more than** that amount. They asked **30** of their teammates for their estimated maximum lift on the bench press exercise. The data ranged from 205 pounds to 385 pounds. The actual different weights were (frequencies are in parentheses) 205(3) 215(3) 225(1) 241(2) 252(2) 265(2) 275(2) 313(2) 316(5) 338(2) 341(1) 345(2) 368(2) 385(1). (Source: data from Reuben Davis, Kraig Evans, and Scott Gunderson.)

Conduct a hypothesis test using a 2.5% level of significance to determine if the bench press mean is **more than 275 pounds**.

Solution:

Set up the Hypothesis Test:

Since the problem is about a mean weight, this is a **test of a single population mean**.

H_o: $\mu = 275$ H_a: $\mu > 275$ This is a right-tailed test.

Calculating the distribution needed:

Random variable: X = the mean weight, in pounds, lifted by the football players.

Distribution for the test: It is normal because σ is known.

$X \sim N\left(275, \frac{55}{\sqrt{30}}\right)$

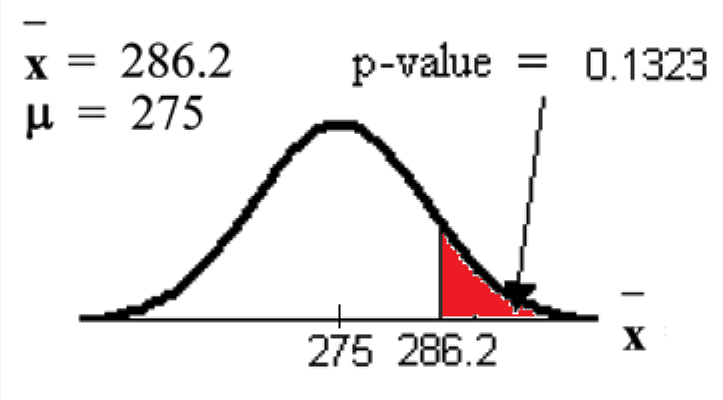
$x = 286.2$ pounds (from the data).

$\sigma = 55$ pounds **(Always use σ if you know it.)** We assume $\mu = 275$ pounds unless our data shows us otherwise.

Calculate the p-value using the normal distribution for a mean and using the sample mean as input (see the calculator instructions below for using the data as input):

p-value $= P(\ x > 286.2) = 0.1323$.

Interpretation of the p-value: If H_o is true, then there is a 0.1331 probability (13.23%) that the football players can lift a mean weight of 286.2 pounds or more. Because a 13.23% chance is large enough, a mean weight lift of 286.2 pounds or more is not a rare event.



Compare α and the p-value:

$\alpha = 0.025$ p-value $= 0.1323$

Make a decision: Since $\alpha <$p-value, do not reject H_o.

Conclusion: At the 2.5% level of significance, from the sample data, there is not sufficient evidence to conclude that the true mean weight lifted is more than 275 pounds.

The p-value can easily be calculated using the TI-83+ and the TI-84 calculators:

Put the data and frequencies into lists. Press `STAT` and arrow over to `TESTS`. Press `1:Z-Test`. Arrow over to `Data` and press `ENTER`.

Arrow down and enter 275 for μ_0, 55 for σ, the name of the list where you put the data, and the name of the list where you put the frequencies. Arrow down to μ : and arrow over to $> \mu_0$. Press `ENTER`. Arrow down to `Calculate` and press `ENTER`. The calculator not only calculates the p-value ($p = 0.1331$, a little different from the above calculation - in it we used the sample mean rounded to one decimal place instead of the data) but it also calculates the test statistic (z-score) for the sample mean, the sample mean, and the sample standard deviation. $\mu > 275$ is the alternate hypothesis. Do this set of instructions again except arrow to `Draw` (instead of `Calculate`). Press `ENTER`. A shaded graph appears with $z = 1.112$ (test statistic) and $p = 0.1331$ (p-value). Make sure when you use `Draw` that no other equations are highlighted in $Y =$ and the plots are turned off.

Example:
Exercise:

Problem:

Statistics students believe that the mean score on the first statistics test is 65. A statistics instructor thinks the mean score is higher than 65. He samples ten statistics students and obtains the scores 65 65 70 67 66 63 63 68 72 71. He performs a hypothesis test using a 5% level of significance. The data are from a normal distribution.

Solution:

Set up the Hypothesis Test:

A 5% level of significance means that $\alpha = 0.05$. This is a test of a **single population mean**.

H_o: $\mu = 65$ $\qquad$ H_a: $\mu > 65$

Since the instructor thinks the average score is higher, use a "$>$". The "$>$" means the test is right-tailed.

Determine the distribution needed:

Random variable: X = average score on the first statistics test.

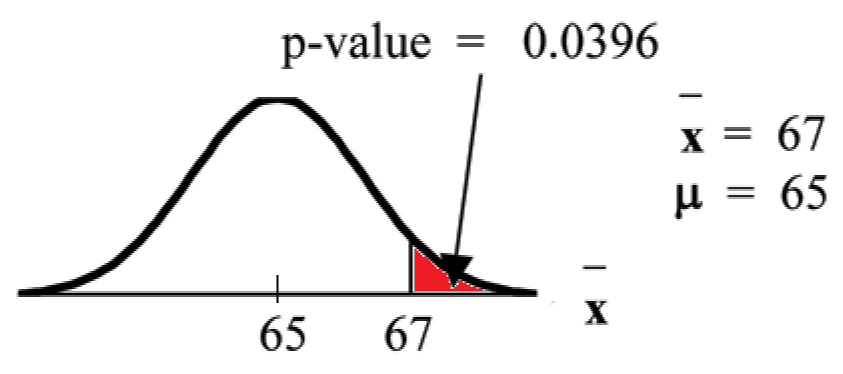
Distribution for the test: If you read the problem carefully, you will notice that there is **no population standard deviation given**. You are only given $n = 10$ sample data values. Notice also that the data come from a normal distribution. This means that the distribution for the test is a student's-t.

Use t_{df}. Therefore, the distribution for the test is t_9 where $n = 10$ and $\text{df} = 10 - 1 = 9$.

Calculate the p-value using the Student's-t distribution:

p-value $= P(\ x > 67\ ) = 0.0396$ where the sample mean and sample standard deviation are calculated as 67 and 3.1972 from the data.

Interpretation of the p-value: If the null hypothesis is true, then there is a 0.0396 probability (3.96%) that the sample mean is 67 or more.



Compare α and the p-value:

Since $\alpha = .05$ and p-value $= 0.0396$. Therefore, $\alpha >$ p-value.

Make a decision: Since $\alpha >$ p-value, reject H_o.

This means you reject $\mu = 65$. In other words, you believe the average test score is more than 65.

Conclusion: At a 5% level of significance, the sample data show sufficient evidence that the mean (average) test score is more than 65, just as the math instructor thinks.

The p-value can easily be calculated using the TI-83+ and the TI-84 calculators:

Put the data into a list. Press `STAT` and arrow over to `TESTS`. Press `2:T-Test`. Arrow over to `Data` and press `ENTER`. Arrow down and enter 65 for μ_0, the name of the list where you put the data, and 1 for `Freq:`. Arrow down to μ : and arrow over to $> \mu_0$. Press `ENTER`. Arrow down to `Calculate` and press `ENTER`. The calculator not only calculates the p-value ($p = 0.0396$) but it also calculates the test statistic (t-score) for the sample mean, the sample mean, and the sample standard deviation. $\mu > 65$ is the alternate hypothesis. Do this set of instructions again except arrow to `Draw` (instead of `Calculate`). Press `ENTER`. A shaded graph appears with $t = 1.9781$ (test statistic) and $p = 0.0396$ (p-value). Make sure when you use `Draw` that no other equations are highlighted in $Y =$ and the plots are turned off.

Example:
Exercise:

Problem:

Joon believes that 50% of first-time brides in the United States are younger than their grooms. She performs a hypothesis test to determine if the percentage is **the same or different from 50%**. Joon samples **100 first-time brides** and **53** reply that they are younger than their grooms. For the hypothesis test, she uses a 1% level of significance.

Solution:

Set up the Hypothesis Test:

The 1% level of significance means that $\alpha = 0.01$. This is a **test of a single population proportion**.

$H_o: p = 0.50$ $\qquad$ $H_a: p \neq 0.50$

The words **"is the same or different from"** tell you this is a two-tailed test.

Calculate the distribution needed:

Random variable: $P\prime$ = the percent of of first-time brides who are younger than their grooms.

Distribution for the test: The problem contains no mention of a mean. The information is given in terms of percentages. Use the distribution for $P\prime$, the estimated proportion.

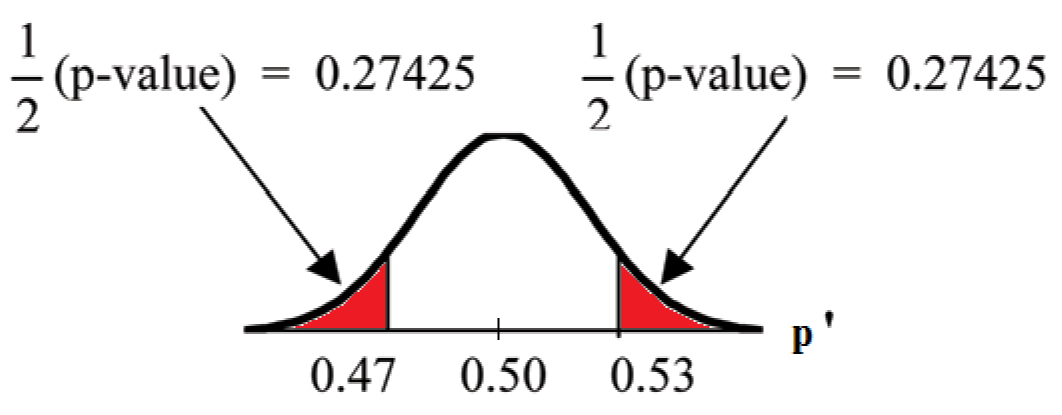
$P\prime \sim N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$ $\qquad$ Therefore, $P\prime \sim N\left(0.5, \sqrt{\frac{0.5 \cdot 0.5}{100}}\right)$ where $p = 0.50$, $q = 1 - p = 0.50$, and $n = 100$.

Calculate the p-value using the normal distribution for proportions:

$\text{p-value} = P(\text{p'} < 0.47 \text{ or p'} > 0.53) = 0.5485$

where $x = 53$, $p\prime = \frac{x}{n} = \frac{53}{100} = 0.53$.

Interpretation of the p-value: If the null hypothesis is true, there is 0.5485 probability (54.85%) that the sample (estimated) proportion $p\prime$ is 0.53 or more OR 0.47 or less (see the graph below).



$\mu = p = 0.50$ comes from H_o, the null hypothesis.

$p\prime = 0.53$. Since the curve is symmetrical and the test is two-tailed, the $p\prime$ for the left tail is equal to $0.50 - 0.03 = 0.47$ where $\mu = p = 0.50$. (0.03 is the difference between 0.53 and 0.50.)

Compare α and the p-value:

Since $\alpha = 0.01$ and p-value $= 0.5485$. Therefore, $\alpha <$ p-value.

Make a decision: Since $\alpha <$p-value, you cannot reject H_o.

Conclusion: At the 1% level of significance, the sample data do not show sufficient evidence that the percentage of first-time brides that are younger than their grooms is different from 50%.

The p-value can easily be calculated using the TI-83+ and the TI-84 calculators:

Press `STAT` and arrow over to `TESTS`. Press `5:1-PropZTest`. Enter .5 for p_0, 53 for x and 100 for n. Arrow down to `Prop` and arrow to `not equals` p_0. Press `ENTER`. Arrow down to

`Calculate` and press `ENTER`. The calculator calculates the p-value ($p = 0.5485$) and the test statistic (z-score). `Prop not equals` .5 is the alternate hypothesis. Do this set of instructions again except arrow to `Draw` (instead of `Calculate`). Press `ENTER`. A shaded graph appears with $z = 0.6$ (test statistic) and $p = 0.5485$ (p-value). Make sure when you use `Draw` that no other equations are highlighted in $Y =$ and the plots are turned off.

The Type I and Type II errors are as follows:

The Type I error is to conclude that the proportion of first-time brides that are younger than their grooms is different from 50% when, in fact, the proportion is actually 50%. (Reject the null hypothesis when the null hypothesis is true).

The Type II error is there is not enough evidence to conclude that the proportion of first time brides that are younger than their grooms differs from 50% when, in fact, the proportion does differ from 50%. (Do not reject the null hypothesis when the null hypothesis is false.)

Example:
Exercise:

Problem:

Suppose a consumer group suspects that the proportion of households that have three cell phones is 30%. A cell phone company has reason to believe that the proportion is 30%. Before they start a big advertising campaign, they conduct a hypothesis test. Their marketing people survey 150 households with the result that 43 of the households have three cell phones.

Solution:

Set up the Hypothesis Test:

$H_o: p = 0.30 \qquad H_a: p \neq 0.30$

Determine the distribution needed:

The **random variable** is P' = proportion of households that have three cell phones.

The **distribution** for the hypothesis test is $P' \sim N\left(0.30, \sqrt{\frac{(0.30)\cdot(0.70)}{150}}\right)$

Exercise:

Problem:

The value that helps determine the p-value is p'. Calculate p'.

Solution:

$p' = \frac{x}{n}$ where x is the number of successes and n is the total number in the sample.

$x = 43, n = 150$

p' = $\frac{43}{150}$

Exercise:

Problem: What is a **success** for this problem?

Solution:

A success is having three cell phones in a household.

Exercise:

Problem: What is the level of significance?

Solution:

The level of significance is the preset α. Since α is not given, assume that $\alpha = 0.05$.

Draw the graph for this problem. Draw the horizontal axis. Label and shade appropriately.

Exercise:

Problem: Calculate the p-value.

Solution:

p-value = 0.7216

Exercise:

Problem:

Make a decision. _____________(Reject/Do not reject) H_0 because_____________.

Solution:

Assuming that $\alpha = 0.05$, $\alpha <$ p-value. The Decision is do not reject H_0 because there is not sufficient evidence to conclude that the proportion of households that have three cell phones is not 30%.

The next example is a poem written by a statistics student named Nicole Hart. The solution to the problem follows the poem. Notice that the hypothesis test is for a single population proportion. This means that the null and alternate hypotheses use the parameter p. The distribution for the test is normal. The estimated proportion $p\prime$ is the proportion of fleas killed to the total fleas found on Fido. This is sample information. The problem gives a preconceived $\alpha = 0.01$, for comparison, and a 95% confidence interval computation. The poem is clever and humorous, so please enjoy it!

Note: Hypothesis testing problems consist of multiple steps. To help you do the problems, solution sheets are provided for your use. Look in the Table of Contents Appendix for the topic "Solution Sheets." If you like, use copies of the appropriate solution sheet for homework problems.

Example:
Exercise:

Problem:

`My dog has so many fleas, They do not come off with ease. As for shampoo, I have tried many types Even one called Bubble Hype, Which only killed 25% of the fleas, Unfortunately I was not pleased. I've used all kinds of soap, Until I had give up hope Until one day I saw An ad that put me in awe. A shampoo used for dogs Called GOOD ENOUGH to Clean a Hog Guaranteed to kill more fleas. I gave Fido a bath And after doing the math His number of fleas Started dropping by 3's! Before his shampoo I counted 42. At the end of his bath, I redid the math And the new shampoo had killed 17 fleas. So now I was pleased. Now it is time for you to have some fun With the level of significance being .01, You must help me figure out Use the new shampoo or go without?`

Solution:

Set up the Hypothesis Test:

$H_{\text{o}}: p = 0.25$ $H_{\text{a}}: p > 0.25$

Determine the distribution needed:

In words, CLEARLY state what your random variable X or P’ represents.

P’ = The proportion of fleas that are killed by the new shampoo

State the distribution to use for the test.

Normal: $N\left(0.25, \sqrt{\frac{(0.25)(1-0.25)}{42}}\right)$

Test Statistic: $z = 2.3163$

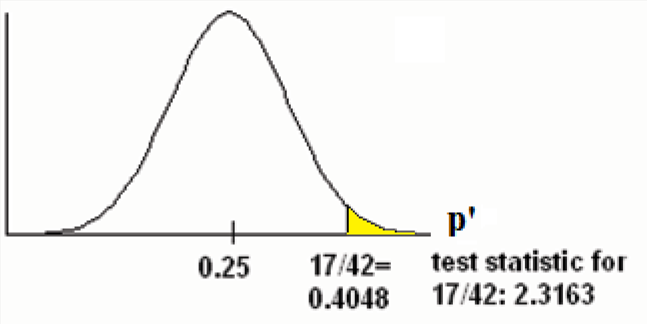
Calculate the p-value using the normal distribution for proportions:

p-value $= 0.0103$

In 1 – 2 complete sentences, explain what the p-value means for this problem.

If the null hypothesis is true (the proportion is 0.25), then there is a 0.0103 probability that the sample (estimated) proportion is 0.4048 $\left(\frac{17}{42}\right)$ or more.

Use the previous information to sketch a picture of this situation. CLEARLY, label and scale the horizontal axis and shade the region(s) corresponding to the p-value.



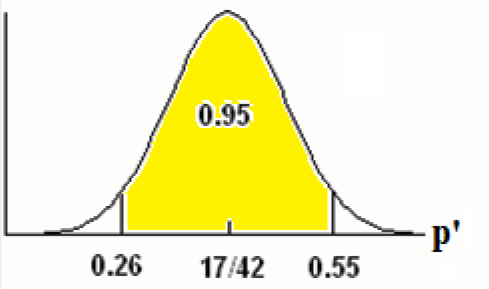
Compare α and the p-value:

Indicate the correct decision (“reject” or “do not reject” the null hypothesis), the reason for it, and write an appropriate conclusion, using COMPLETE SENTENCES.

alpha	decision	reason for decision
0.01	Do not reject H_o	α<p-value

Conclusion: At the 1% level of significance, the sample data do not show sufficient evidence that the percentage of fleas that are killed by the new shampoo is more than 25%.

Construct a 95% Confidence Interval for the true mean or proportion. Include a sketch of the graph of the situation. Label the point estimate and the lower and upper bounds of the Confidence Interval.



Confidence Interval: $(0.26, 0.55)$ We are 95% confident that the true population proportion p of fleas that are killed by the new shampoo is between 26% and 55%.

Note: This test result is not very definitive since the p-value is very close to alpha. In reality, one would probably do more tests by giving the dog another bath after the fleas have had a chance to return.

Glossary

Central Limit Theorem

Given a random variable (RV) with known mean μ and known standard deviation σ. We are sampling with size n and we are interested in two new RVs - the sample mean, $\bar{X}$, and the sample sum, ΣX. If the size n of the sample is sufficiently large, then $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ and $\Sigma X \sim N\left(\mathrm{n\mu}, \sqrt{n}\sigma\right)$. If the size n of the sample is sufficiently large, then the distribution of the sample means and the distribution of the sample sums will approximate a normal distribution regardless of the shape of the population. The mean of the sample means will equal the population mean and the mean of the sample sums will equal n times the population mean. The standard deviation of the distribution of the sample means, $\frac{\sigma}{\sqrt{n}}$, is called the standard error of the mean.

Standard Deviation

A number that is equal to the square root of the variance and measures how far data values are from their mean. Notation: s for sample standard deviation and σ for population standard deviation.

Rational Functions (ticket #5775)
This module will introduce rational functions and describe some of their properties. In particular, it will discuss how rational functions relate to the z-transform and provide a useful tool for characterizing LTI systems.

Introduction

When dealing with operations on polynomials, the term **rational function** is a simple way to describe a particular relationship between two polynomials.

rational function
    For any two polynomials, A and B, their quotient is called a rational function.

Example:
Below is a simple example of a basic rational function, $f(x)$. Note that the numerator and denominator can be polynomials of any order, but the rational function is undefined when the denominator equals zero.
Equation:

$$f(x)=\frac{x^2-4}{2x^2+x-3}$$

If you have begun to study the Z-transform, you should have noticed by now they are all rational functions. Below we will look at some of the properties of rational functions and how they can be used to reveal important characteristics about a z-transform, and thus a signal or LTI system.

Properties of Rational Functions

In order to see what makes rational functions special, let us look at some of their basic properties and characteristics. If you are familiar with rational functions and basic algebraic properties, skip to the next section to see how rational functions are useful when dealing with the z-transform.

Roots

To understand many of the following characteristics of a rational function, one must begin by finding the roots of the rational function. In order to do this, let us factor both of the polynomials so that the roots can be easily determined. Like all polynomials, the roots will provide us with information on many key properties. The function below shows the results of factoring the above rational function, [link].

Equation:

$$f(x)=\frac{(x+2)\,(x-2)}{(2x+3)\,(x-1)}$$

Thus, the roots of the rational function are as follows:

Roots of the numerator are: $\{-2,2\}$

Roots of the denominator are: $\{-3,1\}$

Note: In order to understand rational functions, it is essential to know and understand the roots that make up the rational function.

Discontinuities

Because we are dealing with division of two polynomials, we must be aware of the values of the variable that will cause the denominator of our fraction to be zero. When this happens, the rational function becomes undefined, i.e. we have a discontinuity in the function. Because we have already solved for our roots, it is very easy to see when this occurs. When the variable in the denominator equals any of the roots of the denominator, the function becomes undefined.

Example:
Continuing to look at our rational function above, [link], we can see that the function will have discontinuities at the following points: $x = \{-3, 1\}$

In respect to the Cartesian plane, we say that the discontinuities are the values along the x-axis where the function is undefined. These discontinuities often appear as **vertical asymptotes** on the graph to represent the values where the function is undefined.

Domain

Using the roots that we found above, the **domain** of the rational function can be easily defined.

domain
The group, or set, of values that are defined by a given function.

Example:
Using the rational function above, [link], the domain can be defined as any real number x where x does not equal 1 or negative 3. Written out mathematical, we get the following:
Equation:

$$\{x \in \mathbb{R} |\ (x \neq -3)\ \wedge\ (x \neq 1)\}$$

Intercepts

The **x-intercept** is defined as the point(s) where $f(x)$, i.e. the output of the rational functions, equals zero. Because we have already found the roots of the equation this process is very simple. From algebra, we know that the output will be zero whenever the numerator of the rational function is equal to zero. Therefore, the function will have an x-intercept wherever x equals one of the roots of the numerator.

The **y-intercept** occurs whenever x equals zero. This can be found by setting all the values of x equal to zero and solving the rational function.

Rational Functions and the Z-Transform

As we have stated above, all z-transforms can be written as rational functions, which have become the most common way of representing the z-transform. Because of this, we can use the properties above, especially those of the roots, in order to reveal certain characteristics about the signal or LTI system described by the z-transform.

Below is the general form of the z-transform written as a rational function:
Equation:

$$X(z)=\frac{b_0+b_1z^{-1}+\ldots+b_Mz^{-M}}{a_0+a_1z^{-1}+\ldots+a_Nz^{-N}}$$

If you have already looked at the module about Understanding Pole/Zero Plots and the Z-transform, you should see how the roots of the rational function play an important role in understanding the z-transform. The equation above, [link], can be expressed in factored form just as was done for the simple rational function above, see [link]. Thus, we can easily find the roots of the numerator and denominator of the z-transform. The following two relationships become apparent:

Relationship of Roots to Poles and Zeros

- The roots of the numerator in the rational function will be the **zeros** of the z-transform
- The roots of the denominator in the rational function will be the **poles** of the z-transform

Conclusion

Once we have used our knowledge of rational functions to find its roots, we can manipulate a z-transform in a number of useful ways. We can apply this knowledge to representing an LTI system graphically through a Pole/Zero Plot, or to analyze and design a digital filter through Filter Design from the Z-Transform.

Astronomical Image Deconvolution: Image Processing with Weiner Filter

Our first step was to process the images using the basic procedures used by astronomers in the taking of data. Each set of data images is accompanied by a set of “darks” and “flats” taken at the same time as the rest of teh data. A dark is a picture taken with the telescope closed (i.e. with the lens cap on) for the same length of time as the actual data images are exposed; it produces an “image” of current fluctuations in the CCD due to thermal variation. A flat is an image taken of a white, uniformly lit background, and shows any discrepencies in the CCD produced image, such as dead pixels or positional variations.

Raw Data

[missing_resource: http://cnx.org/GroupWorkspaces/wg614/m14185/rawdata.jpg]

A raw data image, showing errors from the telescope CCD as the lighter left side border.

Calibrated Data

[missing_resource: http://cnx.org/GroupWorkspaces/wg614/m14185/calibrateddata.jpg]

A calibrated data image. Note the removal of the white from the left border.

After we removed the darks and flats from our data images, we were ready to begin processing our calibrated images. For each image, we used the image model we had already made, with a few basic assumptions. First, we assumed an approximately Gaussian distribution for both the noise and the point spread function (PSF) h, which modeled the blurring of the image due to intervening effects, most notably atmospheric disturbance. Secondly, we assumed that the original images of the stars could be modeled as delta functions (point sources) before distortion. Next, we had to build our Weiner Filter G, such that:

Equation:

Deconvolution with Weiner filter G

$$\hat{\mathrm{G}} \times \hat{\mathrm{Y}}\mathrm{j} = \hat{\mathrm{A}}$$

where Â is the Fourier transform of our estimate of the original object, Ĝ is our Weiner filter, and Ŷj is again the Fourier transform of our data image. The Weiner filter is:

Equation:

Addition in a MathML module

$$\frac{\hat{\mathrm{H}} \times \mathrm{Sss}}{\left(|\hat{\mathrm{H}}|^2 + \mathrm{Sss}\right)\mathrm{Snn}} = \hat{\mathrm{G}}$$

with Sss being the power spectrum of the data image, Snn the power spectrum of the noise, and Ĥ the Fourier transform of a Gaussian distribution in two dimensions. However, since our stars are assumed to be point functions, Sss can be assumed as constant; also, we assumed a Gaussian distribution of our noise, so its power spectrum can be modeled by its variance σn2. The Gaussian distribution in two dimensions is given by:

Equation:

Addition in a MathML module

$$(1/(2^*\pi^*\sigma\hat{}2)) \times \exp(\text{-}(u\hat{}2 + v\hat{}2)/(2\sigma\hat{}2)) = h$$

Gaussian Distribution

[missing_resource: http://cnx.org/GroupWorkspaces/wg614/m14185/multiplecornerh.jpg]

Gaussian distribution used for our filter.

where u and v give the position in two dimensions. As seen on the right, our data images take the upper left corner as the origin, so we had to place the center of our Gaussian function there as well as periodize it so we are actually filtering with the entire function across our image. Taking the Fourier transform and substituting in, we got our final Weiner filter. From there, it was a simple matter of multiplying by our data image’s Fourier transform and taking the inverse, two dimensional FFT to find āi, our estimation of the original object from just that one data image. As shown below, the Weiner filter does a good job at sharpening the stars in the image (look in particular at stars in the upper left of the cluster) while not amplifying the noise level. In fact, our measured noise variance remained unchanged through this process.

Processed Image

[missing_resource: http://cnx.org/GroupWorkspaces/wg614/m14185/Try3pic1bestsigma0pt07.jpg]

A figure after the deconvolution process.

Repeating this process for each data image, we created several different estimations of the original object. So, the next major step was to combine these results to get our final best estimate of the object. We began this process by registering the different estimations āi; to do this, we chose three stars on each image as guides. Then, we utilized the image transformation tools available in Matlab to linearly transform (translate/rotate) all three images so that our three guide stars matched up as closely as possible. This step ensured that the images matched up on top of each other for the final step in the process. As the last step in our image processing, we combined the data images using a simple weighted average. Each estimation image of our original object was weighted by the inverse of its original noise σn2. The final product of this weighted average produced our best estimate of the actual Messier object 3.

Form in Music (ticket #5317)
A basic introduction to recognizing form in music.

Form is the Basic Structure

Every piece of music has an overall plan or structure, the "big picture", so to speak. This is called the **form** of the music.

It is easy to recognize and grasp the form of some things, because they are small and simple, like a grain of salt, or repetitive, like a wall made of bricks of the same size. Other forms are easy to understand because they are so familiar; if you see dogs more often than you do sea cucumbers, it should be easier for you to recognize the form of an unfamiliar dog than of an unfamiliar sea cucumber. Other things, like a forest ecosystem, or the structure of a government, are so complex that they have to be explored or studied before their structure can be understood.

Musical forms offer a great range of complexity. Most listeners will quickly grasp the form of a short and simple piece, or of one built from many short repetitions. It is also easier to recognize familiar musical forms. The average American, for example, can distinguish easily between the verses and refrain of any pop song, but will have trouble recognizing what is going on in a piece of music for Balinese gamelan. Classical music traditions around the world tend to encourage longer, more complex forms which may be difficult to recognize without the familiarity that comes from study or repeated hearings.

You can enjoy music without recognizing its form, of course. But understanding the form of a piece helps a musician put together a more credible performance of it. Anyone interested in music theory or history, or in arranging or composing music, must have a firm understanding of form. And being able to "see the big picture" does help the listener enjoy the music even more.

Describing Form

Musicians traditionally have two ways to describe the form of a piece of music. One way involves labelling each large section with a letter. The other way is to simply give a name to a form that is very common.

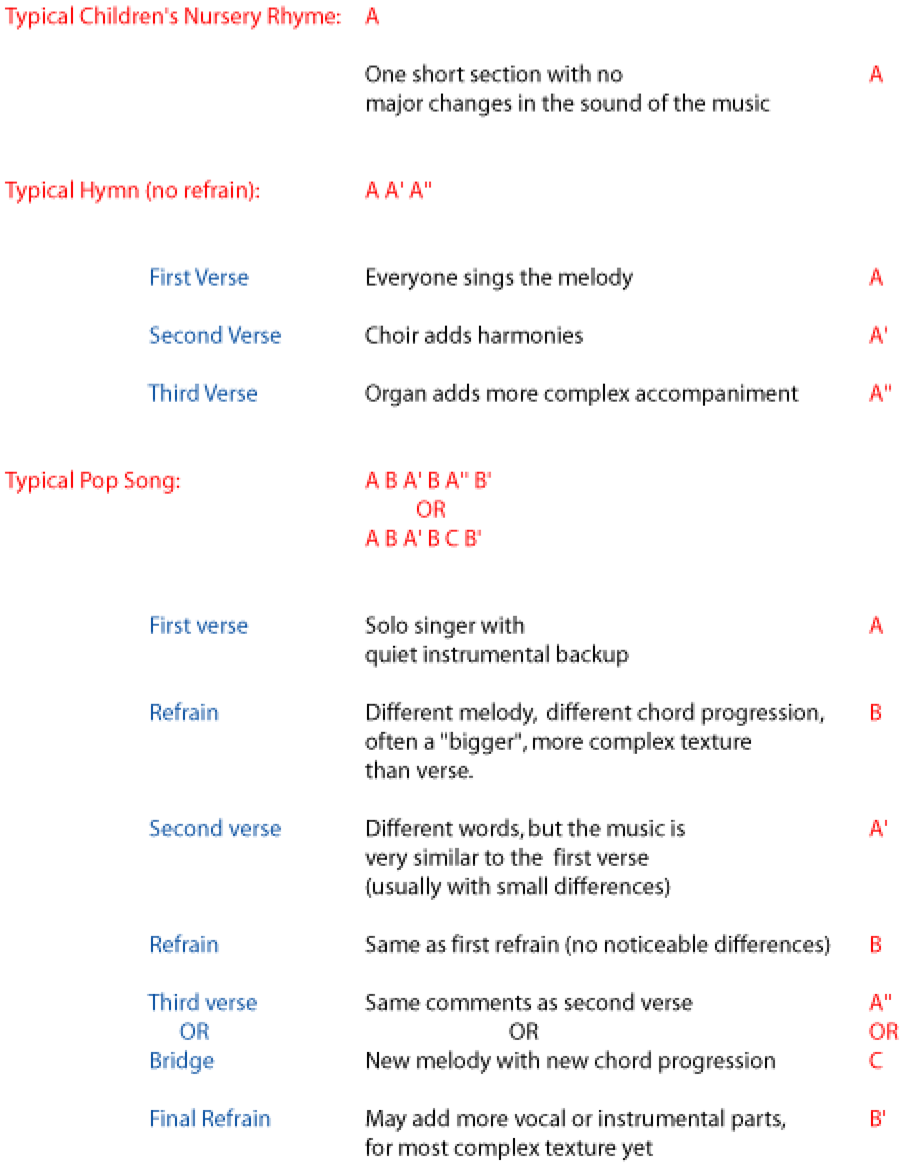
Labelling Form With Letters

Letters can be used to label the form of any piece of music, from the simplest to the most complex. Each major section of the music is labelled with a letter; for example, the first section is the A section. If the second section (or third or fourth) is exactly the same as the first, it is also labelled A. If it is very much like the A section, but with some important differences, it can be labelled A' (pronounced "A prime"). The A' section can also show up later in the piece, or yet another variation of A, A" (pronounced "A double prime") can show up, and so on.

The first major section of the piece that is very different from A is labelled B, and other sections that are like it can be labelled B, B', B", and so on. Sections that are not like A or B are labelled C, and so on.

How do you recognize the sections? With familiar kinds of music, this is pretty easy. (See [link] for some examples of forms that will be familiar to most listeners.) With unfamiliar types of music, it can be more of a challenge. Whether the music is classical, modern, jazz, or pop, listen for repeated sections of music. Also, listen for big changes, in the rhythm, melody, harmony, texture, and timbre. A new section that is not a repetition will usually have noticeable differences in more than one of these areas. For an excellent discussion of form, with plenty of chances to practice hearing the beginnings of new sections, please see Professor Brandt's Sound Reasoning course. In particular, Musical Form deals with recognizing when something new is being introduced (A/B instead of A only), and Time's Effect on the Material deals with recognizing when a section reappears changed (A', B', or A").

Some Familiar Forms



Typical Children's Nursery Rhyme:	A	
	One short section with no major changes in the sound of the music	A
Typical Hymn (no refrain):	A A' A"	
First Verse	Everyone sings the melody	A
Second Verse	Choir adds harmonies	A'
Third Verse	Organ adds more complex accompaniment	A"
Typical Pop Song:	A B A' B A" B' OR A B A' B C B'	
First verse	Solo singer with quiet instrumental backup	A
Refrain	Different melody, different chord progression, often a "bigger", more complex texture than verse.	B
Second verse	Different words, but the music is very similar to the first verse (usually with small differences)	A'
Refrain	Same as first refrain (no noticeable differences)	B
Third verse OR Bridge	Same comments as second verse OR New melody with new chord progression	A" OR C
Final Refrain	May add more vocal or instrumental parts, for most complex texture yet	B'

Most folk and popular music features simple forms that encourage participation.

Exercise:

Problem:

Practice identifying some easy musical forms. Pick some favorite songs and listen to each repeatedly until you are satisfied that you have identified its full form using letters and primes. Compare the forms of the tunes to spot similarities and differences.

Listen for:

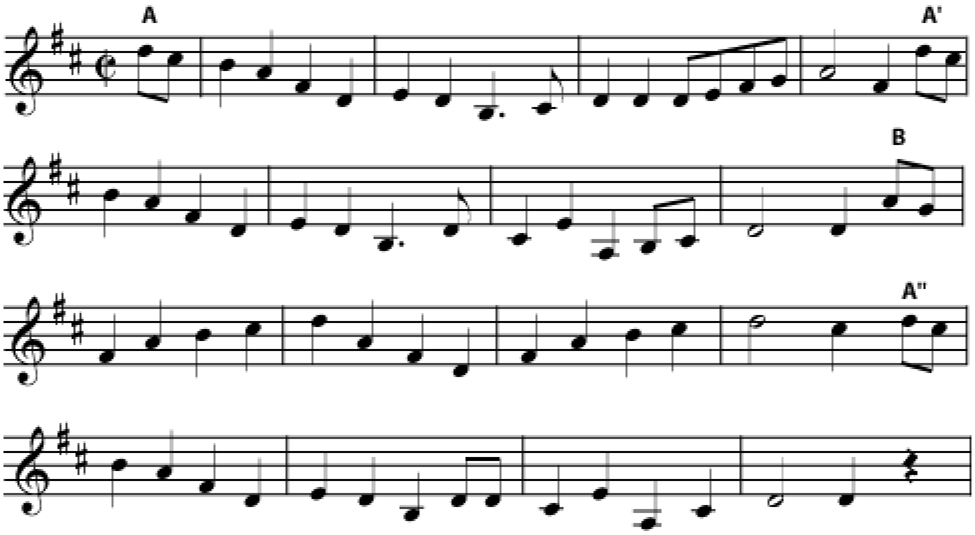
- **Verses** have the same melody but different words.
- **Refrains** have the same melody and the same words.
- **Bridge Sections** are new material that appears late in the song, usually appearing only once or twice, often in place of a verse and usually leading into the refrain. (You may want to note the differences - and the similarity - in the use of the term **bridge** by popular musicians and jazz musicians; see below).
- **Instrumentals** are important sections that have no vocals. They can come at the beginning or end, or in between other sections. Is there more than one? Do they have the same melody as a verse or refrain? Are they similar to each other?

Solution:

Your answers will depend on the songs you choose. Check them with a teacher if you can. (Bring the music so the teacher can listen to it while checking your answers.)

While discussing a piece of music in detail, musicians may also use letters to label smaller parts of the piece within larger sections, even down to labelling individual phrases. For example, the song "The Girl I Left Behind" has many verses with no refrain, an A A' A"- type form. However, a look at the tune of one verse shows that within that overall form is an A A' B A" phrase structure.

Phrase Structure in "The Girl I Left Behind"



In detailed discussions of a piece of music, smaller sections, and even individual phrases, may also be labelled with letters, in order to discuss the piece in greater detail. The A A B A form of this verse is very common, found in verses of everything from folk to jazz to pop music. Verses of blues songs are more likely to have an A A' B form.

Exercise:

Problem:

Now try labeling the phrases of a verse or a refrain of some of the songs you listened to in [link]. Listen for phrases that use similar melodies. (Sometimes, but not always, they even use the same words.) How many of your refrains and verses were basically A A B A? What were the others?

Solution:

If one is available, have a music teacher check your answers.

Naming Forms

Often a musical form becomes so popular with composers that it is given a name. For example, if a piece of music is called a "theme and variations", it is expected to have an overall plan quite different from a piece called a "rondo". (Specifically, the theme and variations would follow an A A' A" A'"... plan, with each section being a new variation on the theme in the first section. A rondo follows an A B A C A ... plan, with a familiar section returning in between sections of new music.)

Also, many genres of music tend to follow a preset form, like the "typical pop song form" in [link]. A **symphony**, for example, is usually a piece of music written for a fairly large number of instruments. It is also associated with a particular form, so knowing that a piece of music is called a symphony should lead you to expect certain things about it. For example, listeners familiar with the symphonic form expect a piece called a symphony to have three or four (depending on when it was written) main sections, called **movements**. They expect a moment of silence in between movements, and also expect the movements to sound very different from each other; for example if the first movement is fast and loud, they might expect that the second movement would be slow and quiet. If they have heard many symphonies, they also would not be at all surprised if the first movement is in sonata form and the third movement is based on a dance.

Note: Although a large group of people who play classical music together is often called a symphony, the more accurate term for the group is **orchestra**. The confusion occurs because many orchestras call themselves "symphony orchestras" because they spend so much time playing symphonies (as opposed to, for example, an "opera orchestra" or a "pops orchestra").

Other kinds of music are also so likely to follow a particular overall plan that they have become associated with a particular form. You can hear musicians talk about something being concerto form or sonata form, for example (even if the piece is not technically a concerto or sonata). Particular dances (a minuet, for example), besides having a set tempo and time signature, will sometimes have a set form that suits the dance steps. And many marches are similar enough in form that there are names for the expected sections (first strain, second strain, trio, break strain).

But it is important to remember that forms are not sets of rules that composers are required to follow. Some symphonies don't have silence between movements, and some don't use the sonata form in any of their movements. Plenty of marches have been written that don't have a trio section, and the development section of a sonata movement can take unexpected turns. And hybrid forms, like the sonata rondo, can become popular with some composers. After all, in architecture, "house" form suggests to most Americans a front and back door, a dining room off the kitchen, and bedrooms with closets, but an architect is free to leave out the dining room, and put the main door at the side of the house and the closets in the bathrooms. Whether a piece of music is a march, a sonata, or a theme and variations, the composer is always free to experiment with the overall architecture of the piece.

Being able to spot that overall architecture as we listen - knowing, so to speak, which room we are in right now - gives us important clues that help us understand and appreciate the music.

Some Common Forms

- **Through-composed** - One section (usually not very long) that does not contain any large repetitions. If a short piece includes repeated phrases, it may be classified by the structure of its phrases.
- **Strophic** - Composed of verses. The music is repeated sections with fairly small changes. May or may not include a refrain.
- **Variations** - One section repeated many times. Most commonly, the melody remains recognizable in each section, and the underlying harmonic structure remains basically the same, but big changes in rhythm, tempo, texture, or timbre keep each section sounding fresh and interesting. Writing a set of variations is considered an excellent exercise for students interested in composing, arranging, and orchestration.

- **Jazz standard song form** - Jazz utilizes many different forms, but one very common form is closely related to the strophic and variation forms. A chord progression in A A B A form (with the B section called the bridge) is repeated many times. On the first and last repetition, the melody is played or sung, and soloists improvise during the other repetitions. The overall form of verse-like repetition, with the melody played only the first and final times, and improvisations on the other repetitions, is very common in jazz even when the A A B A song form is not being used.
- **Rondo** - One section returns repeatedly, with a section of new music before each return. (A B A C A ; sometimes A B A C A B A)
- **Dance forms** - Dance forms usually consist of repeated sections (so there is plenty of music to dance to), with each section containing a set number of measures (often four, eight, sixteen, or thirty-two) that fits the dance steps. Some very structured dance forms (Minuet, for example) are associated even with particular phrase structures and harmonic progressions within each section.
- **Binary Form** - Two different main sections (A B). Commonly in Western classical music, the A section will move away from the tonic, with a strong cadence in another key, and the B section will move back and end strongly in the tonic.
- **Ternary Form** - Three main sections, usually A B A or A B A'.
- **Cyclic Form** - There are two very different uses of this term. One refers to long multimovement works (a "song cycle", for example) that have an overarching theme and structure binding them together. It may also refer to a single movement or piece of music with a form based on the constant repetition of a single short section. This may be an exact repetition (**ostinato**) in one part of the music (for example, the bass line, or the rhythm section), while development, variation, or new melodies occur in other parts. Or it may be a repetition that gradually changes and evolves. This intense-repetition type of cyclic form is very common in folk musics around the world and often finds its way into classical and popular musics, too.
- **Sonata form** - may also be called sonata-allegro or first-movement form. It is in fact often found in the first movement of a sonata, but it has been an extremely popular form with many well-known composers, and so can be found anywhere from the first movement of

a quartet to the final movement of a symphony. In this relatively complex form (too complex to outline here), repetition and development of melodic themes within a framework of expected key changes allow the composer to create a long movement that is unified enough that it makes sense to the listener, but varied enough that it does not get boring.

Autocorrelation of Random Processes

Before diving into a more complex statistical analysis of random signals and processes, let us quickly review the idea of correlation. Recall that the correlation of two signals or variables is the expected value of the product of those two variables. Since our focus will be to discover more about a random process, a collection of random signals, then imagine us dealing with two samples of a random process, where each sample is taken at a different point in time. Also recall that the key property of these random processes is that they are now functions of time; imagine them as a collection of signals. The expected value of the product of these two variables (or samples) will now depend on how quickly they change in regards to **time**. For example, if the two variables are taken from almost the same time period, then we should expect them to have a high correlation. We will now look at a correlation function that relates a pair of random variables from the same process to the time separations between them, where the argument to this correlation function will be the time difference. For the correlation of signals from two different random process, look at the crosscorrelation function.

Autocorrelation Function

The first of these correlation functions we will discuss is the **autocorrelation**, where each of the random variables we will deal with come from the same random process.

Autocorrelation

> the expected value of the product of a random variable or signal realization with a time-shifted version of itself

With a simple calculation and analysis of the autocorrelation function, we can discover a few important characteristics about our random process. These include:

1. How quickly our random signal or processes changes with respect to the time function

2. Whether our process has a periodic component and what the expected frequency might be

As was mentioned above, the autocorrelation function is simply the expected value of a product. Assume we have a pair of random variables from the same process, $X_1 = X(t_1)$ and $X_2 = X(t_2)$, then the autocorrelation is often written as

Equation:

$$\begin{aligned} R_{\mathrm{xx}}(t_1, t_2) &= E[X_1 X_2] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2) \, \mathrm{d}\, x_2 \, \mathrm{d}\, x_1 \end{aligned}$$

The above equation is valid for stationary and nonstationary random processes. For stationary processes, we can generalize this expression a little further. Given a wide-sense stationary processes, it can be proven that the expected values from our random process will be independent of the origin of our time function. Therefore, we can say that our autocorrelation function will depend on the time difference and not some absolute time. For this discussion, we will let $\tau = t_2 - t_1$, and thus we generalize our autocorrelation expression as

Equation:

$$\begin{aligned} R_{\mathrm{xx}}(t, t+\tau) &= R_{\mathrm{xx}}(\tau) \\ &= E[X(t)X(t+\tau)] \end{aligned}$$

for the continuous-time case. In most DSP course we will be more interested in dealing with real signal sequences, and thus we will want to look at the discrete-time case of the autocorrelation function. The formula below will prove to be more common and useful than [link]:

Equation:

$$R_{\mathrm{xx}}[n, n+m] = \sum_{n=-\infty}^{\infty} x[n] x[n+m]$$

And again we can generalize the notation for our autocorrelation function as
Equation:

$$\begin{aligned} R_{\mathrm{xx}}[n, n+m] &= R_{\mathrm{xx}}[m] \\ &= E[X[n]X[n+m]] \end{aligned}$$

Properties of Autocorrelation

Below we will look at several properties of the autocorrelation function that hold for **stationary** random processes.

- Autocorrelation is an even function for τ

$$R_{\mathrm{xx}}(\tau) = R_{\mathrm{xx}}(-\tau)$$

- The mean-square value can be found by evaluating the autocorrelation where $\tau = 0$, which gives us

$$R_{\mathrm{xx}}(0) = X^2$$

- The autocorrelation function will have its largest value when $\tau = 0$. This value can appear again, for example in a periodic function at the values of the equivalent periodic points, but will never be exceeded.

$$R_{\mathrm{xx}}(0) \geq |R_{\mathrm{xx}}(\tau)|$$

- If we take the autocorrelation of a period function, then $R_{\mathrm{xx}}(\tau)$ will also be periodic with the same frequency.

Estimating the Autocorrleation with Time-Averaging

Sometimes the whole random process is not available to us. In these cases, we would still like to be able to find out some of the characteristics of the stationary random process, even if we just have part of one sample function. In order to do this we can **estimate** the autocorrelation from a given interval, 0 to T seconds, of the sample function.

Equation:

$$\check{R}_{\mathrm{xx}}(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} x(t)x(t+\tau)\ \mathrm{d}\ t$$

However, a lot of times we will not have sufficient information to build a complete continuous-time function of one of our random signals for the above analysis. If this is the case, we can treat the information we do know about the function as a discrete signal and use the discrete-time formula for estimating the autocorrelation.

Equation:

$$\check{R}_{\mathrm{xx}}[m] = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x[n]x[n+m]$$

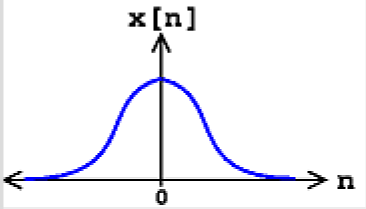
Examples

Below we will look at a variety of examples that use the autocorrelation function. We will begin with a simple example dealing with Gaussian White Noise (GWN) and a few basic statistical properties that will prove very useful in these and future calculations.

Example:

We will let $x[n]$ represent our GWN. For this problem, it is important to remember the following fact about the mean of a GWN function:

$$E[x[n]] = 0$$



Gaussian density function. By examination, can easily see that the above statement is true - the mean equals zero.

Along with being **zero-mean**, recall that GWN is always **independent**. With these two facts, we are now ready to do the short calculations required to find the autocorrelation.

$$R_{\mathrm{xx}}[n, n+m] = E[x[n]x[n+m]]$$

Since the function, $x[n]$, is independent, then we can take the product of the individual expected values of both functions.

$$R_{\mathrm{xx}}[n, n+m] = E[x[n]]E[x[n+m]]$$

Now, looking at the above equation we see that we can break it up further into two conditions: one when m and n are equal and one when they are not equal. When they are equal we can combine the expected values. We are left with the following piecewise function to solve:

$$R_{\mathrm{xx}}[n, n+m] = \begin{cases} E[x[n]]E[x[n+m]] & \text{if } m \neq 0 \\ E\left[x^2[n]\right] & \text{if } m = 0 \end{cases}$$

We can now solve the two parts of the above equation. The first equation is easy to solve as we have already stated that the expected value of $x[n]$ will be zero. For the second part, you should recall from statistics that the expected value of the square of a function is equal to the variance. Thus we get the following results for the autocorrelation:

$$R_{\mathrm{xx}}[n, n+m] = \begin{cases} 0 \ \text{ if } \ m \neq 0 \\ \sigma^2 \ \text{ if } \ m = 0 \end{cases}$$

Or in a more concise way, we can represent the results as

$$R_{\mathrm{xx}}[n, n+m] = \sigma^2 \delta[m]$$

State-Variable Representation of Discrete-Time Systems

State and the State-Variable Representation

State

the minimum additional information at time n, which, along with all current and future input values, is necessary to compute all future outputs.

Essentially, the state of a system is the information held in the delay registers in a filter structure or signal flow graph.

Note: Any LTI (linear, time-invariant) system of finite order M can be represented by a state-variable description

$$\boldsymbol{x}(n+1) = A\boldsymbol{x}(n) + \boldsymbol{B}u(n)$$

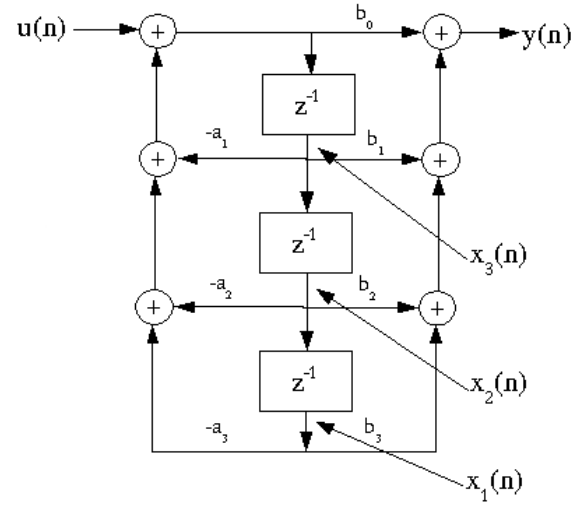
$$\boldsymbol{y}(n) = \boldsymbol{C}\boldsymbol{x}(n) + Du(n)$$

where $\boldsymbol{x}$ is an M x 1 "state vector," $u(n)$ is the input at time n, $y(n)$ is the output at time n; A is an M x M matrix, $\boldsymbol{B}$ is an M x 1 vector, $\boldsymbol{C}$ is a 1 x M vector, and D is a 1 x 1 scalar.

One can always obtain a state-variable description of a signal flow graph.

Example:
3rd-Order IIR

$$y(n) = (-(a_1 y(n-1))) - a_2 y(n-2) - a_3 y(n-3) + b_0 x(n) + b_1 x(n-1) + b_2 x(n-2) + b_3 x(n-3)$$



$$\begin{matrix} x_1(n+1) \\ x_2(n+1) \\ x_3(n+1) \end{matrix} = \begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{matrix} \begin{matrix} x_1(n) \\ x_2(n) \\ x_3(n) \end{matrix} + \begin{matrix} 0 \\ 0 \\ 1 \end{matrix} u(n)$$

$$y(n) = \begin{pmatrix} -(a_3 b_0) & -(a_2 b_0) & -(a_1 b_0) \end{pmatrix} \begin{matrix} x_1(n) \\ x_2(n) \\ x_3(n) \end{matrix} + (b_0)u(n)$$

Exercise:

Problem: Is the state-variable description of a filter $H(z)$ unique?

Exercise:

Problem: Does the state-variable description fully describe the signal flow graph?

State-Variable Transformation

Suppose we wish to define a new set of state variables, related to the old set by a linear transformation: $\boldsymbol{q}(n) = T\boldsymbol{x}(n)$, where T is a nonsingular M x M matrix, and $\boldsymbol{q}(n)$ is the new state vector. We wish the overall system to remain the same. Note that $\boldsymbol{x}(n) = T^{-1}\boldsymbol{q}(n)$, and thus

$$\boldsymbol{x}(n+1) = A\boldsymbol{x}(n) + Bu(n) \Rightarrow T^{-1}\boldsymbol{q}(n) = AT^{-1}\boldsymbol{q}(n) + Bu(n) \Rightarrow \boldsymbol{q}(n) = TAT^{-1}\boldsymbol{q}(n) + TBu(n)$$

$$y(n) = C\boldsymbol{x}(n) + Du(n) \Rightarrow y(n) = CT^{-1}\boldsymbol{q}(n) + Du(n)$$

This defines a new state system with an input-output behavior identical to the old system, but with different internal memory contents (states) and state matrices.

$$\boldsymbol{q}(n) = \hat{A}\boldsymbol{q}(n) + \hat{B}u(n)$$

$$y(n) = \hat{C}\boldsymbol{q}(n) + \hat{D}u(n)$$

$\hat{A} = TAT^{-1}$, $\hat{B} = TB$, $\hat{C} = CT^{-1}$, $\hat{D} = D$

These transformations can be used to generate a wide variety of alternative stuctures or implementations of a filter.

Transfer Function and the State-Variable Description

Taking the z transform of the state equations

$$Z[\boldsymbol{x}(n+1)] = Z[A\boldsymbol{x}(n) + Bu(n)]$$

$$Z[y(n)] = Z[C\boldsymbol{x}(n) + Du(n)]$$

$$\Downarrow$$

$$z\boldsymbol{X}(z) = A\boldsymbol{X}(z) + BU(z)$$

Note: $\boldsymbol{X}(z)$ is a vector of scalar z-transforms $\boldsymbol{X}(z)^T = \begin{pmatrix} X_1(z) & X_2(z) & \dots \end{pmatrix}$

$$\boldsymbol{Y}(z) = C\boldsymbol{X}(n) + DU(n)$$

$$(zI - A)\boldsymbol{X}(z) = BU(z) \Rightarrow \boldsymbol{X}(z) = (zI - A)^{-1}BU(z)$$

so
Equation:

$$\begin{array}{rcl} Y(z) & = & C(zI - A)^{-1}BU(z) + DU(z) \\ & = & \Big(C(-(zI))^{-1}B + D\Big)U(z) \end{array}$$

and thus

$$H(z) = C(zI - A)^{-1}B + D$$

Note that since $(zI - A)^{-1} = \frac{\pm(\det\,(zI - A)_{\text{red}})^{T}}{\det(z(I) - A)}$, this transfer function is an Mth-order rational fraction in z. The denominator polynomial is $D(z) = \det\,(zI - A)$. A discrete-time state system is thus stable if the M roots of $\det\,(zI - A)$ (i.e., the poles of the digital filter) are all inside the unit circle.

Consider the transformed state system with $\hat{A} = TAT^{-1}$, $\hat{B} = TB$, $\hat{C} = CT^{-1}$, $\hat{D} = D$:
Equation:

$$\begin{array}{rcl} H(z) & = & \hat{C}\Big(zI - \hat{A}\Big)^{-1}\hat{B} + \hat{D} \\ & = & CT^{-1}\big(zI - TAT^{-1}\big)^{-1}TB + D \\ & = & CT^{-1}\big(T\,(zI - A)T^{-1}\big)^{-1}TB + D \\ & = & CT^{-1}\big(T^{-1}\big)^{-1}(zI - A)^{-1}T^{-1}TB + D \\ & = & C(zI - A)^{-1}B + D \end{array}$$

This proves that state-variable transformation doesn't change the transfer function of the underlying system. However, it can provide alternate forms that are less sensitive to coefficient quantization or easier to analyze, understand, or implement.

State-variable descriptions of systems are useful because they provide a fairly general tool for analyzing all systems; they provide a more detailed description of a signal flow graph than does the transfer function (although not a full description); and they suggest a large class of alternative implementations. They are even more useful in control theory, which is largely based on state descriptions of systems.

Introduction

Contents of Sampling chapter

- Introduction(Current module)
- Proof
- Illustrations
- Matlab Example
- Hold operation
- System view
- Aliasing applet
- Exercises
- Table of formulas

Why sample?

This section introduces sampling. Sampling is the necessary fundament for all digital signal processing and communication. Sampling can be defined as the process of measuring an analog signal at distinct points.

Digital representation of analog signals offers advantages in terms of

- robustness towards noise, meaning we can send more bits/s
- use of flexible processing equipment, in particular the computer
- more reliable processing equipment
- easier to adapt complex algorithms

Claude E. Shannon

Claude
Elwood
Shannon
(1916-2001)

Claude Shannon has been called the father of information theory, mainly due to his landmark papers on the "Mathematical theory of communication". Harry Nyquist was the first to state the sampling theorem in 1928, but it was not proven until Shannon proved it 21 years later in the paper "Communications in the presence of noise".

Notation

In this chapter we will be using the following notation

- Original analog signal $x(t)$
- Sampling frequency F_s
- Sampling interval T_s (Note that: $F_s = \frac{1}{T_s}$)
- Sampled signal $x_s(n)$. (Note that $x_s(n) = x(nT_s)$)
- Real angular frequency Ω
- Digital angular frequency ω. (Note that: $\omega = \Omega T_s$)

The Sampling Theorem

Note: When sampling an analog signal the sampling frequency must be greater than twice the highest frequency component of the analog signal to be able to reconstruct the original signal from the sampled version.

Finished? Have at look at: Proof Illustrations Matlab Example Aliasing applet Hold operation System view Exercises

Fast Convolution

Fast Circular Convolution

Since,

$$\sum_{m=0}^{N-1} x(m)h(n-m) \bmod N = y(n) \text{is equivalent to} Y(k) = X(k)H(k)$$

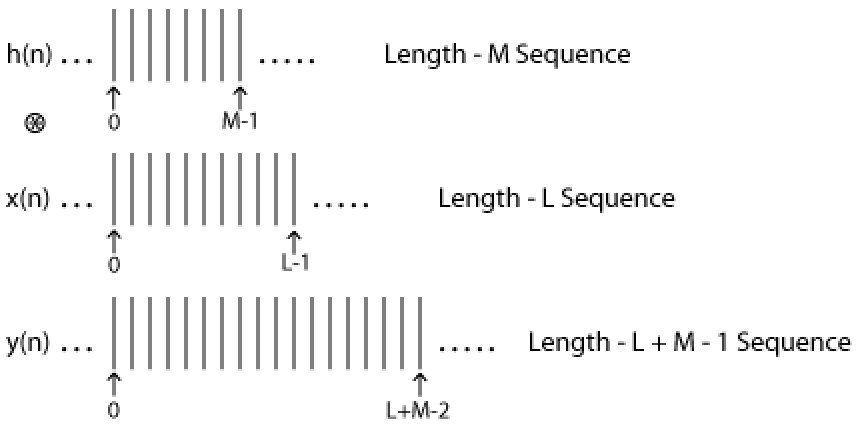
$y(n)$ can be computed as $y(n) = \mathrm{IDFT}\left[\mathrm{DFT}\left[x(n)\right]\mathrm{DFT}\left[h(n)\right]\right]$
Cost

- **Direct**
  - N^2 complex multiplies.
  - $N(N-1)$ complex adds.
- **Via FFTs**
  - 3 FFTs + N multipies.
  - $N + \frac{3N}{2}\log_2 N$ complex multiplies.
  - $3(N\log_2 N)$ complex adds.

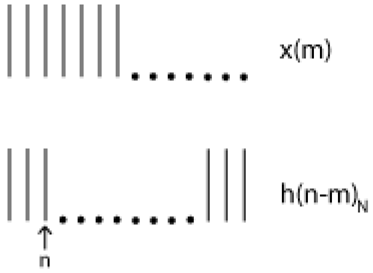
If $H(k)$ can be precomputed, cost is only 2 FFts + N multiplies.

Fast Linear Convolution

DFT produces cicular convolution. For linear convolution, we must zero-pad sequences so that circular wrap-around always wraps over zeros.



To achieve linear convolution using fast circular convolution, we must use zero-padded DFTs of length $N \geq L + M - 1$



Choose shortest convenient N (usually smallest power-of-two greater than or equal to $L + M - 1$)

$$y(n) = \mathrm{IDFT}_N \left[\mathrm{DFT}_N \left[x(n) \right] \mathrm{DFT}_N \left[h(n) \right] \right]$$

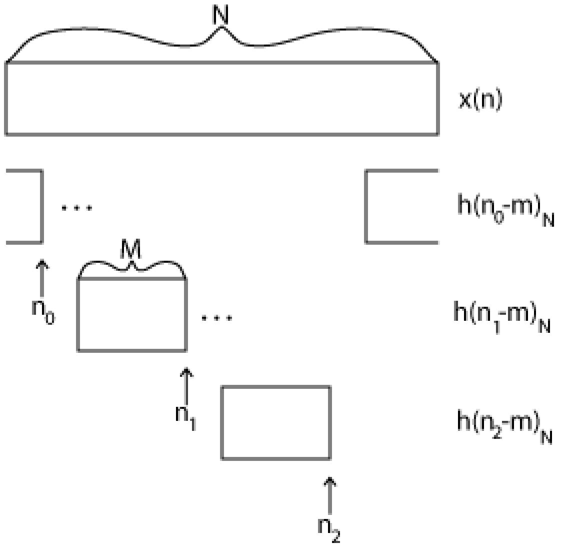
Note: There is some inefficiency when compared to circular convolution due to longer zero-padded DFTs. Still, $O\left(\frac{N}{\log_2 N}\right)$ savings over direct computation.

Running Convolution

Suppose $L = \infty$, as in a real time filter application, or $L \gg M$. There are efficient block methods for computing fast convolution.

Overlap-Save (OLS) Method

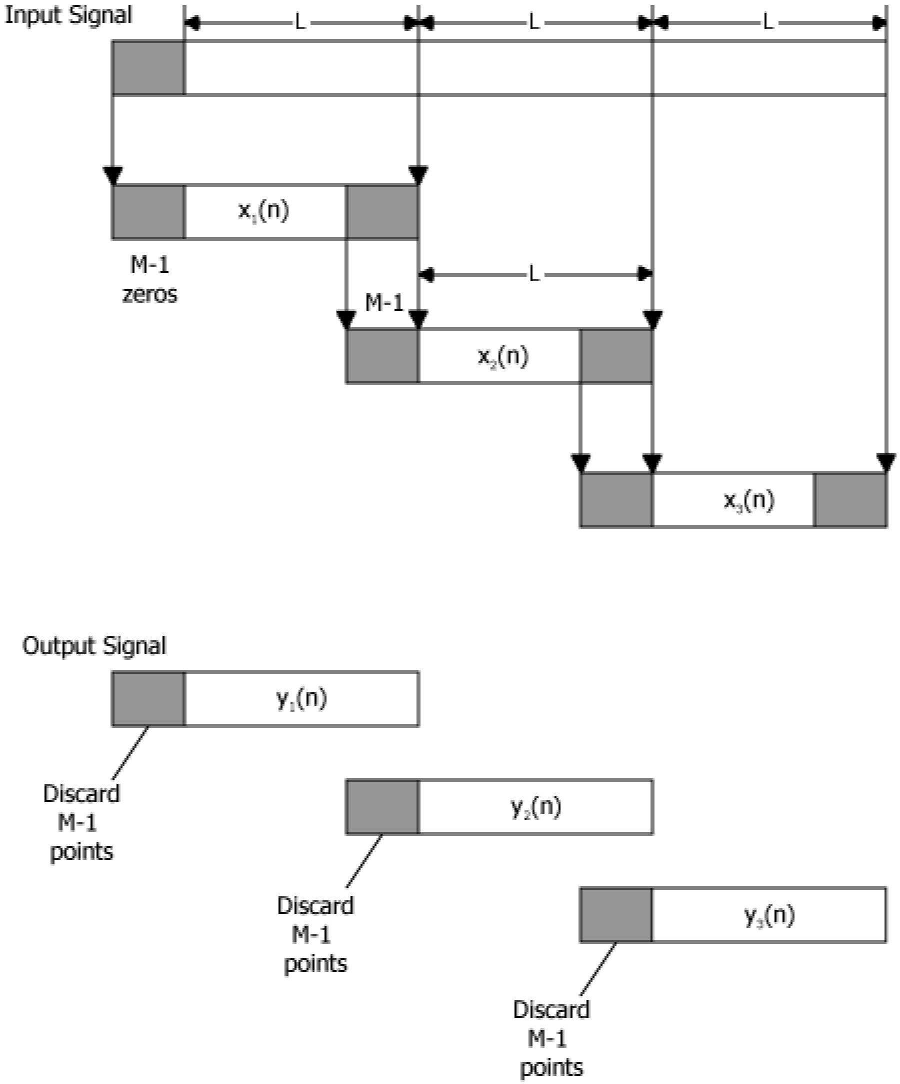
Note that if a length-M filter $h(n)$ is circularly convulved with a length-N segment of a signal $x(n)$,



the first $M - 1$ samples are wrapped around and thus is **incorrect**. However, for $M - 1 \leq n \leq N - 1$,the convolution is linear convolution, so these samples are correct. Thus $N - M + 1$ good outputs are produced for each length-N circular convolution.

The Overlap-Save Method: Break long signal into successive blocks of N samples, each block overlapping the previous block by $M - 1$ samples. Perform circular convolution of each block with filter $h(m)$. Discard first

$M-1$ points in each output block, and concatenate the remaining points to create $y(n)$.



Computation cost for a length-N equals 2^n FFT per output sample is (assuming precomputed $H(k)$) 2 FFTs and N multiplies

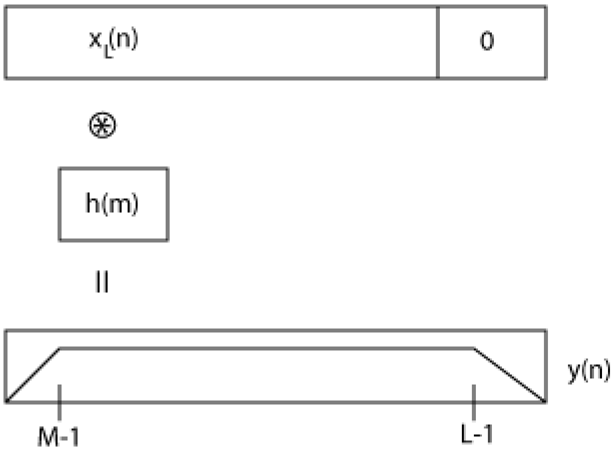
$$\frac{2\left(\frac{N}{2}\log_2 N\right)+N}{N-M+1}=\frac{N\left(\log_2 N+1\right)}{N-M+1}\text{complex multiplies}$$

$$\frac{2\left(N\log_2 N\right)}{N-M+1}=\frac{2N\log_2 N}{N-M+1}\text{complex adds}$$

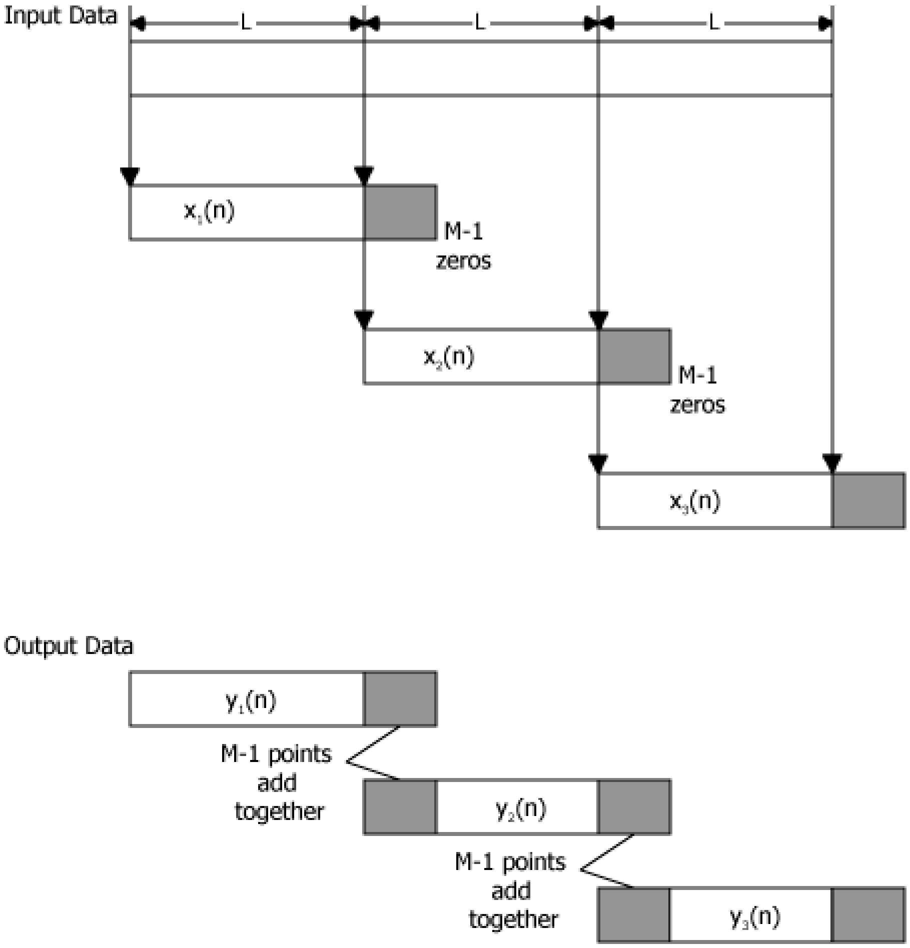
Compare to M mults, $M-1$ adds per output point for direct method. For a given M, optimal N can be determined by finding N minimizing operation counts. Usualy, optimal N is $4M\leq N_{\text{opt}}\leq 8M$.

Overlap-Add (OLA) Method

Zero-pad length-L blocks by $M-1$ samples.



Add successive blocks, overlapped by $M-1$ samples, so that the tails sum to produce the complete linear convolution.



Computational Cost: Two length $N = L + M - 1$ FFTs and M mults and $M - 1$ adds per L output points; essentially the sames as OLS method.

Acquiring a Signal VI

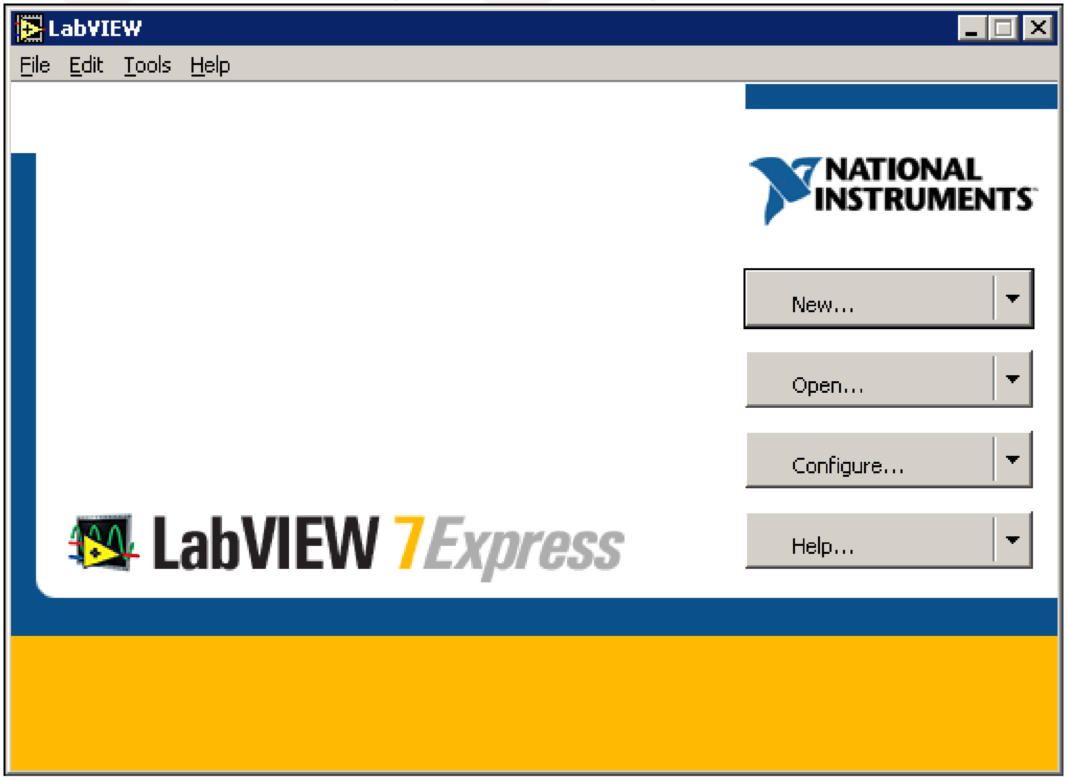
Exercise:

Problem:

In the following exercise, you will build a VI that generates a signal and displays that signal in a graph. LabVIEW provides templates containing information from which you can build a VI. These templates help you get started with LabVIEW.

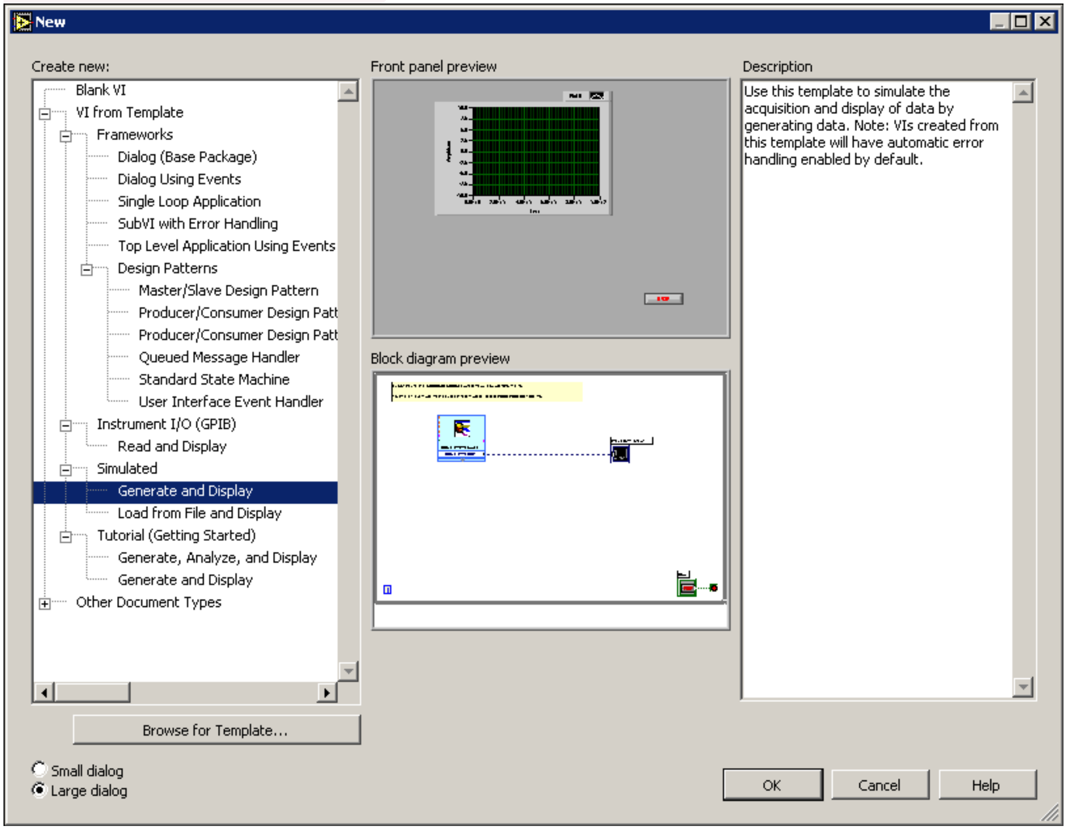
Complete the following steps to create a VI that generates a signal and displays it on the front panel.

1. Launch LabVIEW.
2. In the `LabVIEW` dialog box that appears, shown in [link], click the `New` button to display the `New` dialog box.



3. Select `VI from Template>>Tutorial (Getting Started)>>Generate and Display` in the `Create new` list. This template VI generates and displays a signal. Notice that previews of the template VI appear in the `Front panel preview` and the `Block diagram preview` sections. The

[link] shows the `New` dialog box and the `Generate and Display template` VI.



Use this template to simulate the acquisition and display of data by generating data. Note: VIs created from this template will have automatic error handling enabled by default.

4. Click the `OK` button to open the template. You also can double-click the name of the template VI in the `Create new` list to open the template.
5. Examine the front panel of the VI. The user interface, or front panel, appears with a gray background and includes controls and indicators. The title bar of the front panel indicates that this window is the front panel for the `Generate and Display` VI.

Note: If the front panel is not visible, you can display the front panel by selecting `Window>>Show Front Panel`.

6. Examine the block diagram of the VI. The block diagram appears with a white background and includes VIs and structures that control the front panel objects. The title bar of the block diagram indicates that this window is the block diagram for the `Generate and Display` VI.

   > **Note:** If the block diagram is not visible, you can display the block diagram by selecting `Window>>Show Block Diagram`.

7. ⇨

   On the front panel toolbar, click the `Run` button, shown at left. Notice that a sine wave appears on the graph.

8. STOP

   Stop the VI by clicking the `Stop` button, shown at left, on the front panel.

Adding a Control to the Front Panel

Controls on the front panel simulate the input devices on a physical instrument and supply data to the block diagram of the VI. Many physical instruments have knobs you can turn to change an input value. Complete the following steps to add a knob control to the front panel.

> **Note:** Throughout these exercises, you can undo recent edits by selecting

`Edit>>Undo`

or pressing the

`Ctrl-Z`

keys.

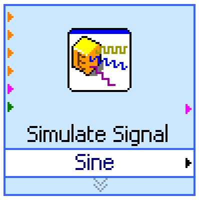
1. If the `Controls` palette is not visible on the front panel, select `Window>>Show Controls Palette` to display it.
2. Move the cursor over the icons on the `Controls` palette to locate the `Numeric Controls` palette. Notice that when you move the cursor over icons on the `Controls` palette, the name of that subpalette appears in the gray space above all the icons on the palette. When you idle the cursor over any icon on any palette, the full name of the subpalette, control, or indicator appears.
3. Click the `Numeric Controls` icon to access the `Numeric Controls` palette.
4. Select the knob control on the `Numeric Controls` palette and place it on the front panel to the left of the waveform graph. You will use this knob in a later exercise to control the amplitude of a signal.
5. Select `File>>Save As` and save this VI as `Acquiring a Signal.vi` in the `C:\Exercises\LabVIEW Basics I` directory.

Note: Save all the VIs you edit or create in this course in the `C:\Exercises\LabVIEW Basics I` directory.

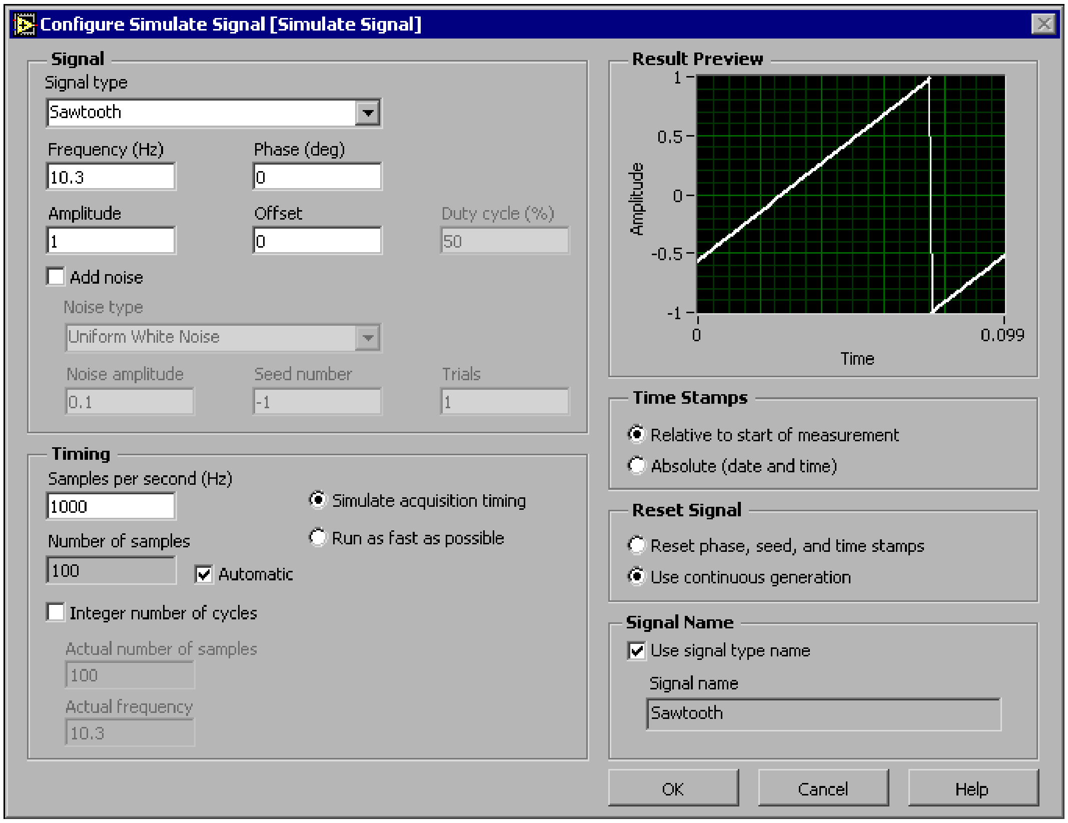
Changing the Signal Type

The block diagram has a blue icon labeled `Simulate Signal`. This icon represents the `Simulate Signal` Express VI. The `Simulate Signal` Express VI simulates a sine wave by default. Complete the following steps to change this signal to a sawtooth wave.

1. Display the block diagram by selecting `Window>>Show Block Diagram` or by clicking the block diagram.

   Notice the `Simulate Signal` Express VI, shown in [link]. An Express VI is a component of the block diagram that you can configure to perform common measurement tasks. The `Simulate Signal` Express VI simulates a signal based on the configuration that you specify.
2. Right-click the `Simulate Signal` Express VI and select `Properties` from the shortcut menu to display the `Configure Simulate Signal` dialog box.
3. Select `Sawtooth` from the `Signal type` pull-down menu. Notice that the waveform on the graph in the `Result Preview` section changes to a sawtooth wave. The `Configure Simulate Signal` dialog box should appear similar to [link].



Signal

Signal type: Sawtooth

Frequency (Hz): 10.3

Phase (deg): 0

Amplitude: 1

Offset: 0

Duty cycle (%): 50

☐ Add noise

Noise type: Uniform White Noise

Noise amplitude: 0.1

Seed number: -1

Trials: 1

Timing

Samples per second (Hz): 1000

● Simulate acquisition timing

○ Run as fast as possible

Number of samples: 100 ☑ Automatic

☐ Integer number of cycles

Actual number of samples: 100

Actual frequency: 10.3

Time Stamps

● Relative to start of measurement

○ Absolute (date and time)

Reset Signal

○ Reset phase, seed, and time stamps

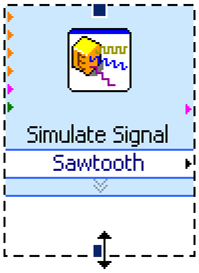
● Use continuous generation

Signal Name

☑ Use signal type name

Signal name: Sawtooth

4. Click the `OK` button to apply the current configuration and close the `Configure Simulate Signal` dialog box.
5. Move the cursor over the down arrows at the bottom of the `Simulate Signal` Express VI.
6.

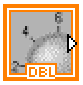


When a double-headed arrow appears, shown at left, click and drag the border of the Express VI until the `Amplitude` input appears. Notice how you expanded the `Simulate Signal` Express VI to display a new input. Because the `Amplitude` input appears on the block diagram, you can configure the amplitude of the sawtooth wave on the block diagram. In [link], notice how `Amplitude` is an option in the `Configure`

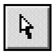
`Simulate Signal` dialog box. When inputs, such as `Amplitude`, appear on the block diagram and in the configuration dialog box, you can configure the inputs in either location.

Wiring Objects on the Block Diagram

To use the knob control to change the amplitude of the signal, you must connect the two objects on the block diagram. Complete the following steps to wire the knob to the `Amplitude` input on the `Simulate Signal` Express VI.

1. 

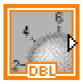
   Move the cursor over the `Knob` terminal, shown at left, until the `Positioning` tool appears.

   Notice how the cursor becomes an arrow, or the `Positioning` tool, shown in [link]. Use the `Positioning` tool to select, position, and resize objects.

2.

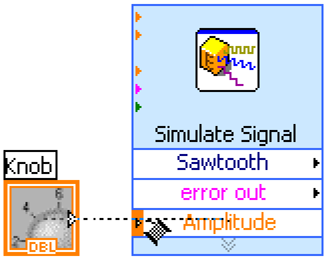
   Click the `Knob` terminal to select it, then drag the terminal to the left of the `Simulate Signal` Express VI. Make sure the `Knob` terminal is inside the loop, shown in [link]. The terminals are representations of front panel controls and indicators. Terminals are entry and exit ports that exchange information between the front panel and block diagram.

3. Deselect the `Knob` terminal by clicking a blank space on the block diagram.
4. 

   Move the cursor over the arrow of the `Knob` terminal, shown in [link].

   Notice how the cursor becomes a wire spool, or the `Wiring` tool, shown at left. Use the `Wiring` tool to wire objects together on the block diagram.

   > **Note:** The cursor does not switch to another tool while an object is selected.

5. 

   When the `Wiring` tool appears, click the arrow and then click the `Amplitude` input of the `Simulate Signal` Express VI, shown in [link], to wire the two objects together. Notice that a wire appears and connects the two objects. Data flows along this wire from the terminal to the Express VI.
6. Select `File>>Save` to save this VI.

Running the VI

Running a VI executes your solution. Complete the following steps to run the `Acquiring a Signal` VI.

1. Display the front panel by selecting `Window>>Show Front Panel` or by clicking the front panel.

   > **Note:** Press the `Ctrl-E` keys to switch from the front panel to the block diagram or from the block diagram to the front panel.

2. Click the `Run` button.
3. Move the cursor over the knob control.

   Notice how the cursor becomes a hand, or the `Operating` tool, shown at left. Use the `Operating` tool to change the value of a control or select the text within a control.
4. Using the `Operating` tool, turn the knob to adjust the amplitude of the sawtooth wave. Notice how the amplitude of the sawtooth wave changes as you turn the knob. Also notice that the y-axis on the graph autoscales to account for the change in amplitude.

   To indicate that the VI is running, the `Run` button changes to a darkened arrow, shown at left. You cannot edit the front panel or block diagram while the VI runs.
5. STOP

   Click the `Stop` button, shown at left, to stop the VI.

Note: Although

`Abort Execution`

button looks like a stop button, the

`Abort Execution`

button does not always properly close the VI. National Instruments recommends stopping your VIs using the

`Stop`

button on the front panel. Use the

`Abort Execution`

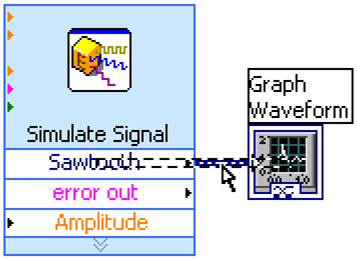
button only when errors prevent you from terminating the application using the

`Stop`

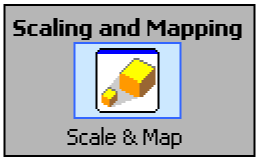
button.

Modifying the Signal

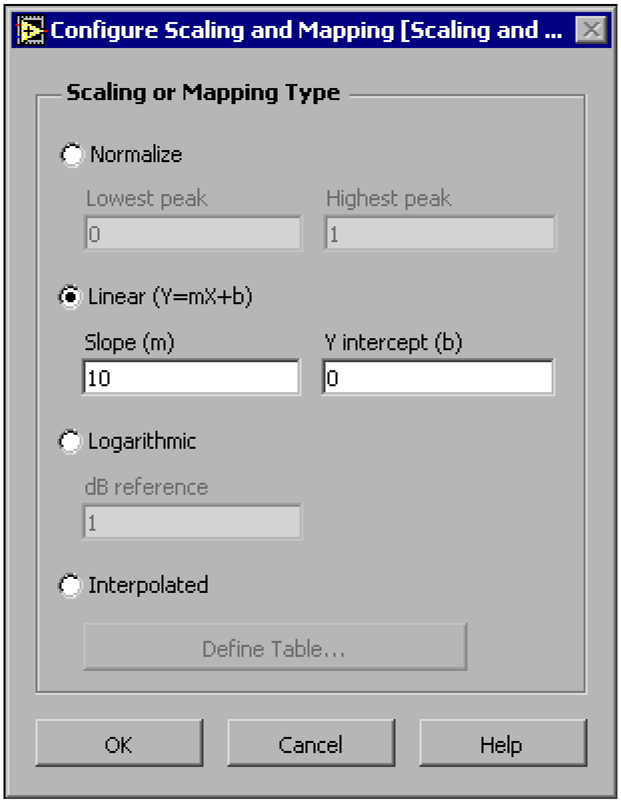
Complete the following steps to add scaling to the signal and display the results in the graph on the front panel.

1. 

On the block diagram, use the `Positioning` tool to double-click the wire that connects the `Simulate Signal` Express VI to the `Waveform Graph` terminal shown in [link].

2. Press the `Delete` key to delete this wire.
3. If the `Functions` palette is not visible on the block diagram, select `Window>>Show Functions Palette` to display it.
4. 

   Select the `Scaling and Mapping` Express VI, shown at left, on the `Arithmetic & Comparison` palette and place it on the block diagram inside the loop between the `Simulate Signal` Express VI and the `Waveform Graph` terminal. If there is no room between the Express VI and the terminal, move the `Waveform Graph` terminal to the right. Notice that the `Configure Scaling and Mapping` dialog box automatically opens when you place the Express VI on the block diagram.
5. Define the value of the scaling factor by entering 10 in the `Slope (m)` text box. The `Configure Scaling and Mapping` dialog box should appear similar to [link].



Configure Scaling and Mapping [Scaling and ...

Scaling or Mapping Type

○ Normalize

Lowest peak: 0

Highest peak: 1

● Linear (Y=mX+b)

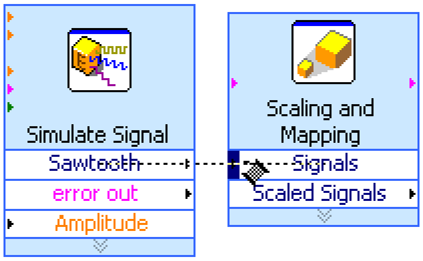
Slope (m): 10

Y intercept (b): 0

○ Logarithmic

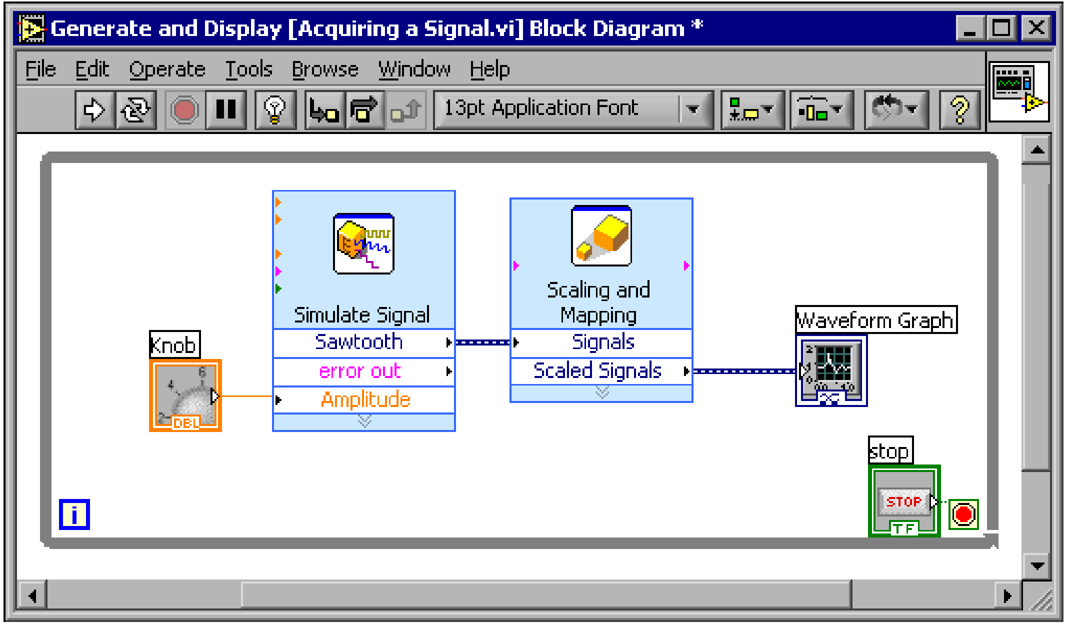
dB reference: 1

○ Interpolated

6. Click the `OK` button to apply the current configuration and close the `Configure Scaling and Mapping` dialog box.
7. Move the cursor over the arrow on the `Sawtooth` output of the `Simulate Signal` Express VI.
8. 

   When the Wiring tool appears, click the arrow and then click the arrow on the `Signals` input of the `Scaling and Mapping` Express VI, shown in [link], to wire the two objects together.
9. Using the `Wiring` tool, wire the `Scaled Signals` output of the `Scaling and Mapping` Express VI to the `Waveform Graph` terminal. Notice the wires connecting the Express VIs and terminals. The arrows on the Express VIs and terminals

indicate the direction that the data flows along these wires. The block diagram should appear similar to [link].



Note: The terminals in the block diagram are set to display as icons. To display a terminal as a data type on the block diagram, right-click the terminal and select `View As Icon` from the shortcut menu to remove the checkmark.

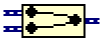
10. Select `File>>Save` to save this VI.

Displaying Two Signals on the Graph

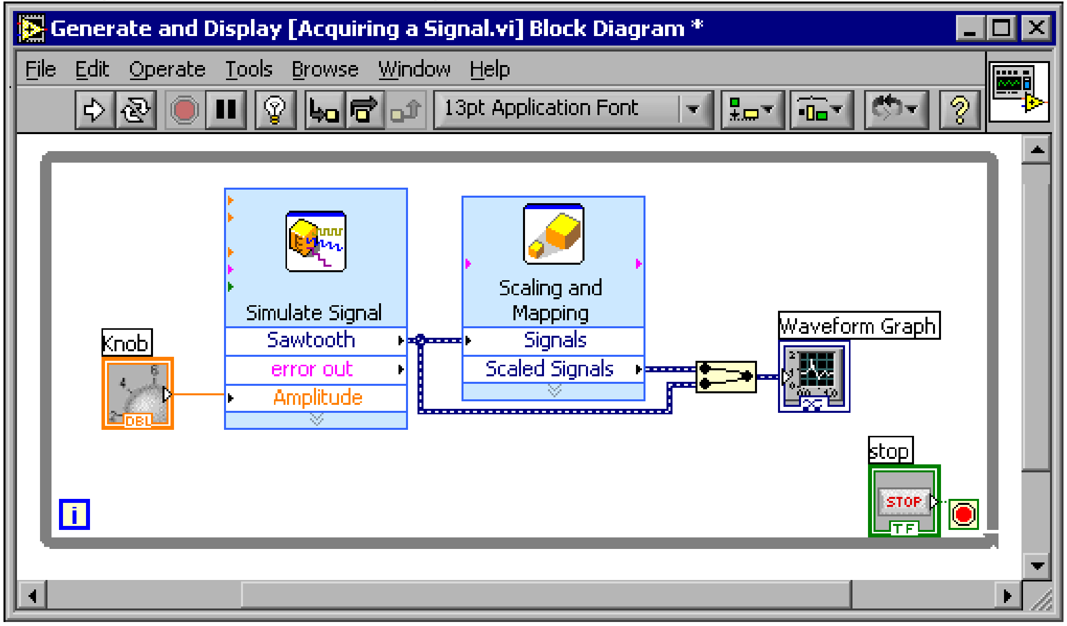
To compare the signal generated by the `Simulate Signal` Express VI and the signal modified by the `Scaling and Mapping` Express VI on the same graph, use the `Merge Signals` function. Complete the following steps to display two signals on the same graph.

1. Move the cursor over the arrow on the `Sawtooth` output of the `Simulate Signal` Express VI.

2. Using the `Wiring` tool, wire the `Sawtooth` output to the `Waveform Graph` terminal.

   The `Merge Signals` function, shown in [link], appears where the two wires connect. This function takes the two separate signals and combines them so that both can be displayed on the same graph. The block diagram should appear similar to [link].

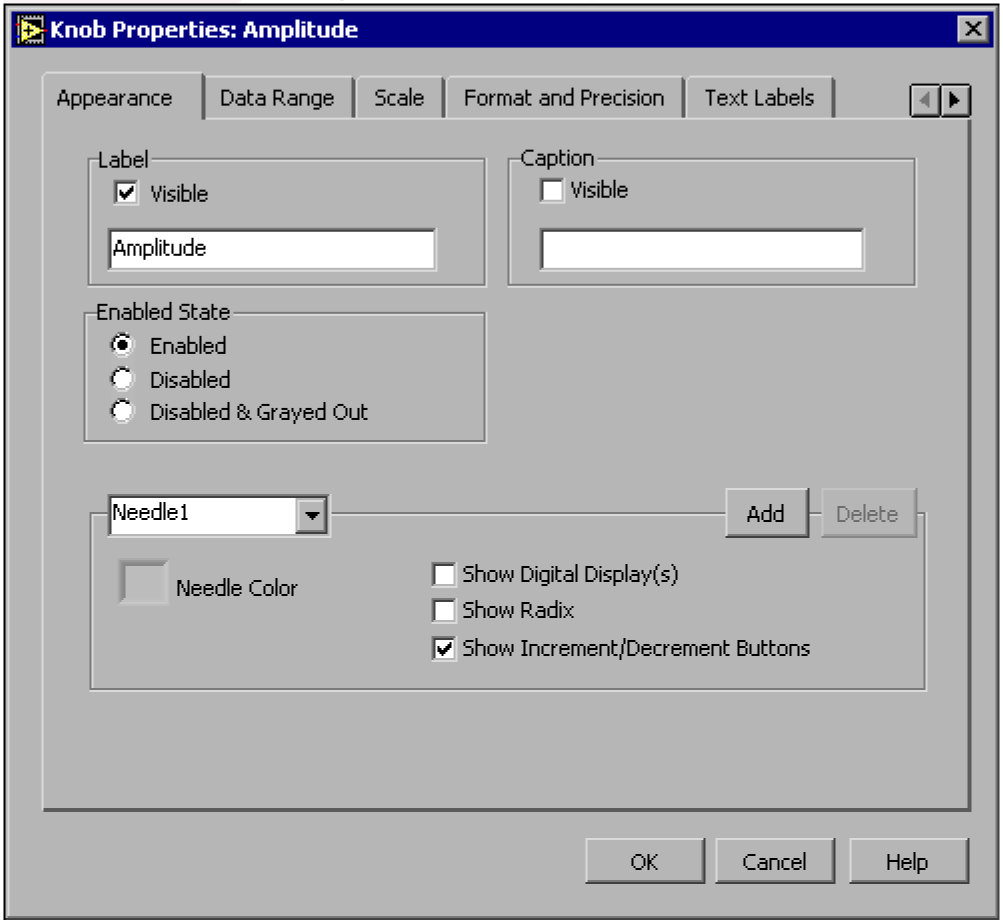
   

3. Select `File>>Save` to save this VI. You also can press the `Ctrl-S` keys to save a VI.
4. Return to the front panel, run the VI, and turn the knob control. Notice that the graph plots the sawtooth wave and the scaled signal. Also notice that the maximum value on the y-axis automatically changes to be 10 times the knob value. This scaling occurs because you set the slope to 10 in the Scaling and Mapping Express VI.
5. Click the `Stop` button.

Customizing the Knob

The knob control changes the amplitude of the sawtooth wave so labeling it `Amplitude` accurately describes the function of the knob. Complete the following steps to customize the appearance of a control on the front panel.

1. Right-click the knob and select `Properties` from the shortcut menu to display the `Knob Properties` dialog box.
2. In the `Label` section on the `Appearance` tab, delete the label Knob, and type `Amplitude` in the text box. The `Knob Properties` dialog box should appear similar to [link].



Label
☑ Visible
Amplitude

Caption
☐ Visible

Enabled State
● Enabled
○ Disabled
○ Disabled & Grayed Out

Needle1
☐ Show Digital Display(s)
☐ Show Radix
☑ Show Increment/Decrement Buttons

3. Click the `Scale` tab and, in the `Scale Range` section, change the maximum value to 5.0. Notice how the knob on the front panel instantly updates to reflect these changes.
4. Click the `OK` button to apply the current configuration and close the `Knob Properties` dialog box.
5. Save this VI.

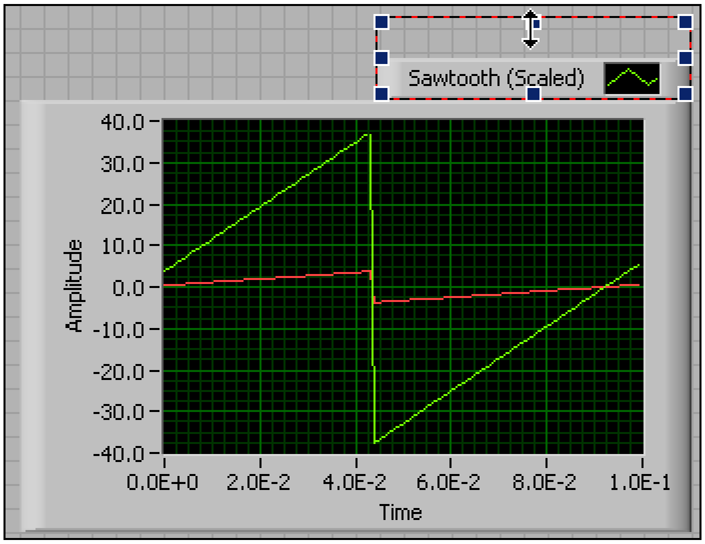
Note: As you build VIs, you can experiment with different properties and configurations. You also can add and delete objects. Remember, you can undo recent edits by selecting `Edit>>Undo` or pressing the `Ctrl-Z` keys.

6. Experiment with other properties of the knob by using the `Knob Properties` dialog box. For example, try changing the colors for the `Marker Text Color` by clicking the color box located on the `Scale` tab.
7. Click the `Cancel` button to avoid applying the changes you made while experimenting. If you want to keep the changes you made, click the `OK` button.

Customizing the Waveform Graph

The `waveform graph` indicator displays the two signals. To indicate which plot is the scaled signal and which is the simulated signal, you customize the plots. Complete the following steps to customize the appearance of an indicator on the front panel.

1. Move the cursor over the top of the `plot legend` on the `waveform graph`. Notice that while there are two plots on the graph, the `plot legend` displays only one plot.
2. When a double-headed arrow appears, shown in [link], click and drag the border of the `plot legend` until the second plot name appears.



3. Right-click the `waveform graph` and select `Properties` from the shortcut menu to display the `Graph Properties` dialog box.
4. On the `Plots` tab, select `Sawtooth` from the pull-down menu. Click the `Line Color` color box to display the `color picker`. Select a new line color.
5. Select `Sawtooth (Scaled)` from the pull-down menu.
6. Place a checkmark in the `Don't use waveform names for plot names` checkbox.
7. In the `Name` text box, delete the current label and change the name of this plot to `Scaled Sawtooth`.
8. Click the `OK` button to apply the current configuration and close the `Graph Properties` dialog box. Notice how the plot color on the front panel changes.
9. Experiment with other properties of the graph by using the `Graph Properties` dialog box. For example, try disabling the autoscale feature located on the `Scales` tab.
10. Click the `Cancel` button to avoid applying the changes you made while experimenting. If you want to keep the changes you made, click the `OK` button.
11. Save and close this VI.

Relative Atomic Masses and Empirical Formulae

Foundation

We begin by assuming the central postulates of the **Atomic-Molecular Theory**. These are: the elements are comprised of identical atomsall atoms of a single element have the same characteristic massthe number and masses of these atoms do not change during a chemical transformation compounds consist of identical molecules formed of atoms combined in simple whole number ratios. We also assume a knowledge of the observed natural laws on which this theory is based: the **Law of Conservation of Mass**, the **Law of Definite Proportions**, and the **Law of Multiple Proportions**.

Goals

We have concluded that atoms combine in simple ratios to form molecules. However, we don't know what those ratios are. In other words, we have not yet determined any molecular formulae. In the second table of Concept Development Study #1, we found that the mass ratios for nitrogen oxide compounds were consistent with many different molecular formulae. A glance back at the nitrogen oxide data shows that the oxide B could be NO, NO_2, N_2O, or any other simple ratio.

Each of these formulae correspond to different possible relative atomic weights for nitrogen and oxygen. Since oxide B has oxygen to nitrogen ratio 1.14 : 1, then the relative masses of oxygen to nitrogen could be 1.14:1 or 2.28:1 or 0.57:1 or many other simple possibilities. If we knew the relative masses of oxygen and nitrogen atoms, we could determine the molecular formula of oxide B. On the other hand, if we knew the molecular formula of oxide B, we could determine the relative masses of oxygen and nitrogen atoms. If we solve one problem, we solve both. Our problem then is that we need a simple way to "count" atoms, at least in relative numbers.

Observation 1: Volume Relationships in Chemical Reactions

Although mass is conserved, most chemical and physical properties are not conserved during a reaction. Volume is one of those properties which is not conserved, particularly when the reaction involves gases as reactants or products. For example, hydrogen and oxygen react explosively to form water vapor. If we take 1 liter of oxygen gas and 2 liters of hydrogen gas, by careful analysis we could find that the reaction of these two volumes is complete, with no left over hydrogen and oxygen, and that 2 liters of water vapor are formed. Note that the total volume is not conserved: 3 liters of oxygen and hydrogen become 2 liters of water vapor. (All of the volumes are measured at the same temperature and pressure.)

More notable is the fact that the ratios of the volumes involved are simple whole number ratios: 1 liter of oxygen : 2 liters of hydrogen : 2 liters of water. This result proves to be general for reactions involving gases. For example, 1 liter of nitrogen gas reacts with 3 liters of hydrogen gas to form 2 liters of ammonia gas. 1 liter of hydrogen gas combines with 1 liter of chlorine gas to form 2 liters of hydrogen chloride gas. These observations can be generalized into the **Law of Combining Volumes**.

Law of Combining Volumes

When gases combine during a chemical reaction at a fixed pressure and temperature, the ratios of their volumes are simple whole number ratios.

These simple integer ratios are striking, particularly when viewed in the light of our conclusions from the Law of Multiple Proportions. Atoms combine in simple whole number ratios, and evidently, volumes of gases also combine in simple whole number ratios. Why would this be? One simple explanation of this similarity would be that the volume ratio and the ratio of atoms and molecules in the reaction are the same. In the case of the hydrogen and oxygen, this would say that the ratio of volumes (1 liter of oxygen : 2 liters of hydrogen : 2 liters of water) is the same as the ratio of atoms and molecules (1 atom of oxygen: 2 atoms of hydrogen: 2 molecules of water). For this to be true, equal volumes of gas would have to contain equal numbers of gas particles (atoms or molecules), independent of the type of gas. If true, this means that the volume of a gas must be a direct measure of the number of particles (atoms or molecules) in the gas. This would allow us to "count" the number of gas particles and determine molecular formulae.

There seem to be big problems with this conclusion, however. Look back at the data for forming hydrogen chloride: 1 liter of hydrogen plus 1 liter of chlorine yields 2 liters of hydrogen chloride. If our thinking is true, then this is equivalent to saying that 1 hydrogen atom plus 1 chlorine atom makes 2 hydrogen chloride molecules. But how could that be possible? How could we make 2 identical molecules from a single chlorine atom and a single hydrogen atom? This would require us to divide each hydrogen and chlorine atom, violating the postulates of the atomic-molecular theory.

Another problem appears when we weigh the gases: 1 liter of oxygen gas weighs more than 1 liter of water vapor. If we assume that these volumes contain equal numbers of particles, then we must conclude that 1 oxygen particle weighs more than 1 water particle. But how could that be possible? It would seem that a water molecule, which contains at least one oxygen atom, should weigh more than a single oxygen particle.

These are serious objections to the idea that equal volumes of gas contain equal numbers of particles. Our postulate appears to have contradicted common sense and experimental observation. However, the simple ratios of the Law of Combining Volumes are also equally compelling. Why should volumes react in simple whole number ratios if they do not represent equal numbers of particles? Consider the opposite viewpoint: if equal volumes of gas do not contain equal numbers of particles, then equal numbers of particles must be contained in unequal volumes not related by integers. Now when we combine particles in simple whole number ratios to form molecules, the volumes of gases required would produce decidedly non-whole number ratios. The Law of Combining Volumes should not be contradicted lightly.

There is only one logical way out. We will accept our deduction from the Law of Combining Volumes that **equal volumes of gas contain equal numbers of particles**, a conclusion known as **Avogadro's Hypothesis**. How do we account for the fact that 1 liter of hydrogen plus 1 liter of chlorine yields 2 liters of hydrogen chloride? There is only one way for a single hydrogen particle to produce 2 identical hydrogen chloride molecules: each hydrogen particle must contain more than one atom. In fact, each hydrogen particle (or molecule) must contain an even number of hydrogen atoms. Similarly, a chlorine molecule must contain an even number of chlorine atoms.

More explicitly, we observe that
Equation:

$$1 \text{ liter of hydrogen} + 1 \text{ liter of chlorine} \rightarrow 2 \text{ liters of hydrogen chloride}$$

Assuming that each liter volume contains an equal number of particles, then we can interpret this observation as
Equation:

$$1H_2\text{molecule} + 1\ \mathrm{Cl}_2\ \text{molecule} \rightarrow 2H\ \mathrm{Cl}\ \text{molecules}$$

(Alternatively, there could be any fixed even number of atoms in each hydrogen molecule and in each chlorine molecule. We will assume the simplest possibility and see if that produces any contradictions.)

This is a wonderful result, for it correctly accounts for the Law of Combining Volumes and eliminates our concerns about creating new atoms. Most importantly, we now know the molecular formula of hydrogen chloride. We have, in effect, found a way of "counting" the atoms in the reaction by measuring the volume of gases which react.

This method works to tell us the molecular formula of many compounds. For example,
Equation:

$$2 \text{ liters of hydrogen} + 1 \text{ liter of oxygen} \rightarrow 2 \text{ liters of water}$$

This requires that oxygen particles contain an even number of oxygen atoms. Now we can interpret this equation as saying that
Equation:

$$2H_2\text{molecules} + 1O_2\text{molecule} \rightarrow 2H_2O\text{molecules}$$

Now that we know the molecular formula of water, we can draw a definite conclusion about the relative masses of the hydrogen and oxygen atoms. Recall from the Table that the mass ratio in water is 8:1 oxygen to hydrogen.

Since there are two hydrogen atoms for every oxygen atom in water, then the mass ratio requires that a single oxygen atom weigh 16 times the mass of a hydrogen atom.

To determine a mass scale for atoms, we simply need to choose a standard. For example, for our purposes here, we will say that a hydrogen atom has a mass of 1 on the atomic mass scale. Then an oxygen atom has a mass of 16 on this scale.

Our conclusions account for the apparent problems with the masses of reacting gases, specifically, that oxygen gas weighs more than water vapor. This seemed to be nonsensical: given that water contains oxygen, it would seem that water should weigh more than oxygen. However, this is now simply understood: a water molecule, containing only a single oxygen atom, has a mass of 18, whereas an oxygen molecule, containing two oxygen atoms, has a mass of 32.

Determination of Atomic Weights for Gaseous Elements

Now that we can count atoms and molecules to determine molecular formulae, we need to determine relative atomic weights for all atoms. We can then use these to determine molecular formulae for any compound from the mass ratios of the elements in the compound.

We begin by examining data on reactions involving the Law of Combining Volumes. Going back to the nitrogen oxide data given here, we recall that there are three compounds formed from nitrogen and oxygen. Now we measure the volumes which combine in forming each. We find that 2 liters of oxide B can be decomposed into 1 liter of nitrogen and 1 liter of oxygen. From the reasoning above, then a nitrogen particle must contain an even number of nitrogen atoms. We assume for now that nitrogen is N_2. We have already concluded that oxygen is O_2. Therefore, the molecular formula for oxide B is NO, and we call it nitric oxide. Since we have already determined that the oxygen to nitrogen mass ratio is 1.14 : 1, then, if we assign oxygen a mass of 16, as above, nitrogen has a mass of 14. (That is $\frac{16}{1.14} = 14$.) 2 liters of oxide A is formed from 2 liters of oxygen and 1 liter of nitrogen. Therefore, oxide A is NO_2, which we call nitrogen dioxide. Note that we predict an oxygen to nitrogen mass ratio of $\frac{32}{14} = 2.28 : 1$, in agreement with the data. Oxide C is N_2O, called nitrous oxide, and predicted to have a mass ratio of $\frac{16}{28} = 0.57 : 1$, again in agreement with the data. We have now resolved the ambiguity in the molecular formulae.

What if nitrogen were actually N_4? Then the first oxide would be N_2O, the second would be N_2O_2, and the third would be N_4O. Furthermore, the mass of a nitrogen atom would be 7. Why don't we assume this? Simply because in doing so, we will always find that the minimum relative mass of nitrogen in any molecule is 14. Although this might be two nitrogen atoms, there is no reason to believe that it is. Therefore, a single nitrogen atom weighs 14, and nitrogen gas particles are N_2.

Determination of Atomic Weights for Non-Gaseous Elements

We can proceed with this type of measurement, deduction, and prediction for any compound which is a gas and which is made up of elements which are gases. But this will not help us with the atomic masses of non-gaseous elements, nor will it permit us to determine the molecular formulae for compounds which contain these elements.

Consider carbon, an important example. There are two oxides of carbon. Oxide A has oxygen to carbon mass ratio 1.33 : 1 and oxide B has mass ratio 2.66 : 1. Measurement of reacting volumes shows that we find that 1 liter of oxide A is produced from 0.5 liters of oxygen. Hence, each molecule of oxide A contains only half as many oxygen atoms as does an oxygen molecule. Oxide A thus contains one oxygen atom. But how many carbon atoms does it contain? We can't determine this yet because the elemental carbon is solid, not gas. This means that we also cannot determine what the mass of a carbon atom is.

But we can try a different approach: we weigh 1 liter of oxide A and 1 liter of oxygen gas. The result we find is that oxide A weighs 0.875 times per liter as much as oxygen gas. Since we have assumed that a fixed volume of gas contains a fixed number of particles, then 1 liter of oxide A contains just as many particles as 1 liter of oxygen gas. Therefore, each **particle** of oxide A weighs 0.875 times as much as a particle of oxygen gas (that is, an O_2

molecule). Since an O_2 molecule weighs 32 on our atomic mass scale, then a particle of oxide A weighs $0.875 \times 32 = 28$. Now we know the molecular weight of oxide A.

Furthermore, we have already determined from the combining volumes that oxide A contains a single oxygen atom, of mass 16. Therefore, the mass of carbon in oxide A is 12. However, at this point, we do not know whether this is one carbon atom of mass 12, two atoms of mass 6, eight atoms of mass 1.5, or one of many other possibilities.

To make further progress, we make additional measurements on other carbon containing gas compounds. 1 liter of oxide B of carbon is formed from 1 liter of oxygen. Therefore, each oxide B molecule contains two oxygen atoms. 1 liter of oxide B weighs 1.375 times as much as 1 liter of oxygen. Therefore, one oxide B molecule has mass $1.375 \times 32 = 44$. Since there are two oxygen atoms in a molecule of oxide B, the mass of oxygen in oxide B is 32. Therefore, the mass of carbon in oxide B is 12, the same as in oxide A.

We can repeat this process for many such gaseous compounds containing carbon atoms. In each case, we find that the mass of carbon in each molecule is either 12 or a multiple of 12. We never find, for examples, 6 or 18, which would be possible if each carbon atom had mass 6. The simplest conclusion is that a carbon atom has mass 12. Once we know the atomic mass of carbon, we can conclude that the molecular formula of oxide A is CO, and that of oxide B is CO_2.

Therefore, the atomic masses of non-gaseous elements can be determined by mass and volume measurements on gaseous compounds containing these elements. This procedure is fairly general, and most atomic masses can be determined in this way.

Moles, Molecular Formulae and Stoichiometric Calculations

We began with a circular dilemma: we could determine molecular formulae provided that we knew atomic masses, but that we could only determine atomic masses from a knowledge of molecular formulae. Since we now have a method for determining all atomic masses, we have resolved this dilemma and we can determine the molecular formula for any compound for which we have percent composition by mass.

As a simple example, we consider a compound which is found to be 40.0% carbon, 53.3% oxygen, and 6.7% hydrogen by mass. Recall from the Law of Definite Proportions that these mass ratios are independent of the sample, so we can take any convenient sample to do our analysis. Assuming that we have 100.0g of the compound, we must have 40.0g of carbon, 53.3g of oxygen, and 6.7g of hydrogen. If we could count or otherwise determine the number of atoms of each element represented by these masses, we would have the molecular formula. However, this would not only be extremely difficult to do but also unnecessary.

From our determination of atomic masses, we can note that 1 atom of carbon has a mass which is 12.0 times the mass of a hydrogen atom. Therefore, the mass of N atoms of carbon is also 12.0 times the mass of N atoms of hydrogen atoms, no matter what N is. If we consider this carefully, we discover that 12.0g of carbon contains exactly the same number of atoms as does 1.0g of hydrogen. Similarly, we note that 1 atom of oxygen has a mass which is $\frac{16.0}{12.0}$ times the mass of a carbon atom. Therefore, the mass of N atoms of oxygen is $\frac{16.0}{12.0}$ times the mass of N atoms of carbon. Again, we can conclude that 16.0g of oxygen contains exactly the same number of atoms as 12.0g of carbon, which in turn is the same number of atoms as 1.0g of hydrogen. Without knowing (or necessarily even caring) what the number is, we can say that it is the same number for all three elements.

For convenience, then, we **define** the number of atoms in 12.0g of carbon to be 1 **mole** of atoms. Note that 1 mole is a specific number of particles, just like 1 dozen is a specific number, independent of what objects we are counting. The advantage to defining the mole in this way is that it is easy to determine the number of moles of a substance we have, and knowing the number of moles is equivalent to counting the number of atoms (or molecules) in a sample. For example, 24.0g of carbon contains 2.0 moles of atoms, 30.0g of carbon contains 2.5 moles of atoms, and in general, x grams of carbon contains $\frac{x}{12.0}$ moles of atoms. Also, we recall that 16.0g of oxygen contains exactly as many atoms as does 12.0g of carbon, and therefore 16.0g of oxygen contains exactly 1.0 mole of oxygen atoms. Thus, 32.0g of oxygen contains 2.0 moles of oxygen atoms, 40.0g of oxygen contains 2.5 moles, and x grams of oxygen contains $\frac{x}{16.0}$ moles of oxygen atoms. Even more generally, then, if we have m grams of an element whose atomic mass is M, the number of moles of atoms, n, is

Equation:

$$n = \frac{m}{M}$$

Now we can determine the relative numbers of atoms of carbon, oxygen, and hydrogen in our unknown compound above. In a 100.0g sample, we have 40.0g of carbon, 53.3g of oxygen, and 6.7g of hydrogen. The number of moles of atoms in each element is thus

Equation:

$$\begin{aligned} n_C &= \frac{40.0g}{12.0\frac{g}{\text{mol}}} \\ &= 3.33\,\text{moles} \end{aligned}$$

Equation:

$$\begin{aligned} n_O &= \frac{53.3g}{16.0\frac{g}{\text{mol}}} \\ &= 3.33\,\text{moles} \end{aligned}$$

Equation:

$$\begin{aligned} n_H &= \frac{6.7g}{1.0\frac{g}{\text{mol}}} \\ &= 6.67\,\text{moles} \end{aligned}$$

We note that the numbers of moles of atoms of the elements are in the simple ratio $n_C : n_O : n_H = 1 : 1 : 2$. Since the number of particles in 1 mole is the same for all elements, then it must also be true that the number of atoms of the elements are in the simple ratio 1 : 1 : 2. Therefore, the molecular formula of the compound must be COH_2.

Or is it? On further reflection, we must realize that the simple ratio 1 : 1 : 2 need not represent the exact numbers of atoms of each type in a molecule of the compound, since it is indeed only a ratio. Thus the molecular formula could just as easily be $C_2O_2H_4$ or $C_3O_3H_6$. Since the formula COH_2 is based on empirical mass ratio data, we refer to this as the **empirical formula** of the compound. To determine the **molecular formula**, we need to determine the relative mass of a molecule of the compound, i.e. the molecular mass. One way to do so is based on the Law of Combining Volumes, Avogadro's Hypothesis, and the **Ideal Gas Law**. To illustrate, however, if we were to find that the relative mass of one molecule of the compound is 60.0, we could conclude that the molecular formula is $C_2O_2H_4$.

Review and Discussion Questions

Exercise:

Problem:

State the Law of Combining Volumes and provide an example of your own construction which demonstrates this law.

Exercise:

Problem:

Explain how the Law of Combining Volumes, combined with the Atomic-Molecular Theory, leads directly to Avogadro's Hypothesis that equal volumes of gas at equal temperatures and pressure contain equal numbers of particles.

Exercise:

Problem:

Use Avogadro's Hypothesis to demonstrate that oxygen gas molecules cannot be monatomic.

Exercise:

Problem:

The density of water vapor at room temperature and atmospheric pressure is $0.737\frac{g}{L}$. Compound A is 80.0% carbon by mass, and 20.0% hydrogen. Compound B is 83.3% carbon by mass and 16.7% hydrogen. The density of gaseous Compound A is $1.227\frac{g}{L}$, and the density of Compound B is $2.948\frac{g}{L}$. Show how these data can be used to determine the molar masses of Compounds A and B, assuming that water has molecular mass 18.

Exercise:

Problem:

From the results above, determine the mass of carbon in a molecule of Compound A and in a molecule of Compound B. Explain how these results indicate that a carbon atom has atomic mass 12.

Exercise:

Problem:

Explain the utility of calculating the number of moles in a sample of a substance.

Exercise:

Problem:

Explain how we can conclude that 28g of nitrogen gas (N_2) contains exactly as many molecules as 32g of oxygen gas (O_2), even though we cannot possibly count this number.